



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

**PENERAPAN MODIFIED K-NEAREST NEIGHBOR (MKNN)
UNTUK KLASIFIKASI DATA SERANGAN JARINGAN
KOMPUTER (NSL KDD CUP 1999)**

TUGAS AKHIR

**Diajukan Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik Pada
Jurusan Teknik Informatika**

Oleh:

FAKHRIAL IRSYADI

11351101913



FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU

PEKANBARU

2019



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERSETUJUAN

PENERAPAN MODIFIED K-NEAREST NEIGHBOR (MKNN) UNTUK KLASIFIKASI DATA SERANGAN JARINGAN KOMPUTER (NSL KDD CUP 1999)

TUGAS AKHIR

Oleh

FAKHRIAL IRSYADI
11351101913

Telah diperiksa dan disetujui sebagai laporan tugas akhir
di Pekanbaru, pada tanggal 18 Desember 2019

Pembimbing,

Iwan Iskandar, M.T.
NIP.19821216 201503 1 003

ii



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PENGESAHAN

**PENERAPAN MODIFIED K-NEAREST NEIGHBOR (MKNN)
UNTUK KLASIFIKASI DATA SERANGAN JARINGAN
KOMPUTER (NSL KDD CUP 1999)**

TUGAS AKHIR

Oleh


FAKHRIAL IRSYADI
11351101913

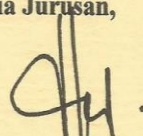
Telah dipertahankan di depan sidang dewan penguji
Sebagai salah satu syarat untuk memperoleh gelar sarjana Teknik Informatika
Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau
di Pekanbaru, pada tanggal 18 Desember 2019

Pekanbaru, 18 Desember 2019

Mengesahkan

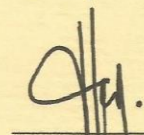
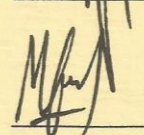
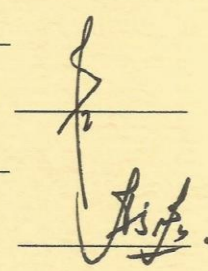
Ketua Jurusan,


Dekan,
Dr. Drs. Ahmad Darmawi, M.Ag.
NIP. 19660604 199203 1 004


Dr. Elin Haerani, S.T., M.Kom.
NIP. 19810523 200710 2 003

DEWAN PENGUJI

Ketua : Dr. Elin Haerani, S.T., M.Kom
Sekretaris : Iwan Iskandar, M.T.
Penguji I : Muhammad Fikry, S.T., M.Sc.
Penguji II : Iis Afrianty, S.T., M.Sc.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL

Tugas Akhir yang tidak diterbitkan ini terdaftar dan tersedia di Perpustakaan Universitas Islam Negeri Sultan Syarif Kasim Riau adalah terbuka untuk umum dengan ketentuan bahwa hak cipta pada penulis. Referensi kepustakaan diperkenankan dicatat, tetapi pengutipan atau ringkasan hanya dapat dilakukan dengan izin penulis dan harus disertai dengan kebiasaan ilmiah untuk menyebutkan sumbernya.

Penggandaan atau penerbitan sebagian atau seluruh Tugas Akhir ini harus memperoleh izin dari Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau. Perpustakaan yang meminjamkan Tugas Akhir ini untuk anggotanya diharapkan untuk mengisi nama, tanda peminjaman dan tanggal pinjam.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan didalam daftar pustaka.

Pekanbaru, 18 Desember 2019

Yang membuat pernyataan,

FAKHRIAL IRSYADI

11351101913

UIN SUSKA RIAU



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERSEMBAHAN

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

*Bacalah dengan menyebut nama Tuhanmu
Karena sesungguhnya Sesudah kesulitan itu ada kemudahan
Sesungguhnya sesudah kesulitan itu ada kemudahan
(Qs. Al-Insyirah : 5-6)*

*Alhamdulillahilladzi bi ni'matihi tatimmush shalihati, puji dan syukur kepada Allah
'azza wa jalla Dzat maha mulia dan agung yang berkehendak atas segala sesuatu.*

Atas kemudahan yang diberikan Allah 'azza wa ja jalla kepada ana

Dalam menyelesaikan tugas akhir ini.

*Shalawat serta salam semoga senantiasa tercurahkan kepada baginda Rasulullah
shallallahu 'alaihi wa sallam, kemudian kepada para keluarga beliau, sahabat-sahabat
beliau serta kepada para pengikut nya hingga akhir zaman kelak,*

*Dengan rasa penuh cinta yang mendalam dan ucapan terima kasih yang sangat besar
jazaakumullahu khairan kepada kedua orang tercinta, ayahanda Jamalis Lunin dan
ibunda Ita Chendra.*

*Yang senantiasa memberikan doa serta motivasi kepada ana sehingga ana dapat
melewati segala rintangan dan tantangan dalam menyelesaikan tugas akhir ini.*



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

PENERAPAN MODIFIED K-NEAREST NEIGHBOR (MKNN) UNTUK KLASIFIKASI DATA SERANGAN JARINGAN KOMPUTER (NSL KDD CUP 1999)

FAKHRIAL IRSYADI

11351101913

Jurusan Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Sultan Syarif Kasim Riau

ABSTRAK

Serangan jaringan komputer ialah suatu tindakan yang mengancam keamanan jaringan dengan merusak seluruh atau sebagian sumber daya jaringan pada lingkup tertentu. Terdapat beberapa Jenis serangan komputer diantaranya *port scanning*, *syn attack*, *smurf attack*, *ip spoofing* serta berbagai macam serangan lainnya. Dengan massifnya jenis serangan yang terjadi maka dibutuhkan *machine learning* untuk mengklasifikasi serangan jaringan tersebut. Penelitian ini menggunakan metode *Modified K-Nearest Neighbor* untuk mengklasifikasikan berbagai jenis serangan. data yang digunakan pada penelitian ini ialah *dataset* NSL-KDD yang merupakan versi pembaharuan dari versi sebelumnya yakni dataset KDD cup 1999 yang terdiri dari data latih dan data uji yang telah melewati tahapan proses KDD. Dengan jumlah data latih ialah 88.724 dan data uji 14.537. Dataset ini memiliki 39 jenis serangan dan 1 kelas normal. Jumlah variabel yang digunakan ialah 29 dengan 5 kelas yakni Normal, DoS, R2L, U2R dan probe. Parameter yang digunakan ialah k=3, k=5, k=7, k=9 dan k=11. Dari penelitian ini metode MKNN telah berhasil mengklasifikasikan serangan jaringan pada data NSL-KDD dengan tingkat akurasi tertinggi yakni 86,31% pada nilai k=7.

Kata kunci : *Dataset NSL-KDD, Klasifikasi, Machine Learning, Serangan, Modified K-Nearest Neighbor*

UIN SUSKA RIAU



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

IMPLEMENTATION OF MODIFIED K-NEAREST NEIGHBOR (MKNN) FOR CLASSIFICATION OF COMPUTER NETWORKS AT DATA (NSL KDD CUP 1999)

FAKHRIAL IRSYADI
11351101913

Informatics Engineering Departement
Faculty of Science and Technology
State Islamic University Of Sultan Syarif Kasim Riau

ABSTRACT

A computer network attack is an action that threatens network security by destroying all or part of the network's resources in a certain scope. There are several types of computer attacks such as port scanning, syn attacks, smurf attacks, IP spoofing and various other attacks. With the massive types of attacks that occur, machine learning is needed to classify these network attacks. This study uses the Modified K-Nearest Neighbor method to classify various types of attacks. The data used in this study is the NSL-KDD dataset which is an updated version of the previous version, the KDD Cup 1999 dataset consisting of training data and test data that have passed the stages of the KDD process. the number of training data is 88.724 and the test data is 14.537. This dataset has 39 types of attacks and 1 normal class. The number of variables used is 29 with 5 classes namely Normal, DoS, R2L, U2R and probes. the parameters used are $k = 3$, $k = 5$, $k = 7$, $k = 9$ and $k = 11$. From the research method of MKNN has completely to classify network attack on NSL-KDD data which obtained highest accuracy is 86,31% at $k=7$.

Keywords: *Dataset NSL-KDD, Classification, Machine Learning Attack, Modified K-Nearest Neighbor,*

UIN SUSKA RIAU



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Assalamu 'alaikum Warahmatullahi Wabarakaatuh.

Alhamdulillah Rabbil 'Alamin. Puji dan syukur penulis haturkan kepada Allah 'azza wa jalla atas segala rahmat beserta karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini dengan judul **“Penerapan Modified K-Nearest Neighbor (MKNN) untuk Klasifikasi Data Serangan Jaringan Komputer (NSL KDD CUP 1999)”**. Shalawat serta salam kepada Nabi Muhammad shallallahu 'alaihi wa sallam kemudian kepada para keluarga beliau, sahabat-sahabat beliau dan kepada pengikutnya hingga akhir zaman kelak.

Laporan tugas akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana di Jurusan Teknik Informatika Universitas Islam Negeri Sultan Syarif Kasim Riau. Dalam penyusunan tugas akhir ini penulis banyak mendapatkan bantuan dari banyak pihak. Terutama karena adanya kehendak serta kemudahan yang diberikan oleh Allah 'azza wa jalla penulis dapat menyelesaikan tugas akhir ini. Selain itu penulis ingin mengucapkan rasa terima kasih yang tak terhingga kepada ayahanda tercinta Jamalis Lunin dan ibunda tercinta Ita Chendra yang telah memberikan dukungan penuh kepada penulis dalam penyusunan Tugas Akhir ini. Pada kesempatan ini penulis juga ingin mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. H. Akhmad Mujahidin, M.Ag, selaku Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
2. Bapak Dr. Ahmad Darmawi, M.Ag, selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
3. Ibu Dr. Elin Haerani, S.T., M.Kom. selaku Ketua Jurusan Teknik Informatika, Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau
4. Ibu Siska Kurnia Gusti, S.T., M.Sc. selaku dosen Penasehat Akademik penulis selama kuliah di Universitas Islam Negeri Sultan Syarif Kasim Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

5. Bapak Iwan Iskandar, S.T., M.T. selaku pembimbing Tugas Akhir penulis yang telah banyak memberikan bimbingan, saran serta motivasi kepada penulis hingga penulis dapat menyelesaikan Tugas Akhir ini.
6. Bapak Muhammad Fikry, S.T., M.Sc. dan Ibu Iis Afrianty, S.T., M.Sc. selaku penguji I dan Penguji II yang telah memberikan bimbingan, kritik serta saran dalam pengerjaan tugas akhir ini.
7. Bapak dan Ibu dosen Teknik Informatika yang telah mendidik serta memberikan pengajaran sehingga penulis mendapatkan gelar S1 di Teknik Informatika Universitas Islam Negeri Sultan Syarif Kasim Riau.
8. Terima kasih untuk sahabat-sahabat sesama perjuangan penulis jeprianto, jasriadi hasyir, aldi wiratama, rivalza fahlevi, panji kuswoyo, reza kurnianda serta seluruh kru kos jep yang telah memberikan semangat dalam penyusunan tugas akhir ini.
9. Teman-teman TIF E 2013, terima kasih atas dukungan, motivasi, kekompakan serta semangat yang kalian berikan selama ini
10. Teman-teman angkatan TIF 2013, terima kasih atas dukungan serta semangat yang diberikan selama ini kepada penulis.
11. Serta seluruh pihak yang telah membantu penulis yang tidak bisa disebutkan satu persatu, penulis haturkan ucapan terima kasih.

Mungkin dalam Tugas Akhir ini terdapat banyak kesalahan yang telah penulis lakukan, untuk itu penulis ucapkan mohon maaf sebesar-besarnya, serta pembaca dapat memberi kritik dan saran ke fakhrial.irsyadi@students.uin-suska.ac.id dan fakhrial31@gmail.com dan besar harapan penulis semoga Tugas Akhir ini dapat bermanfaat untuk pembaca terutama untuk penulis. Allahumma Aamiin

Pekanbaru, 18 Desember 2019

FAKHRIAL IRSYADI



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR ISI

	Halaman
LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL	iv
LEMBAR PERNYATAAN	v
LEMBAR PERSEMBAHAN	vi
ABSTRAK	vii
ABSTRACT	viii
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR GAMBAR	xv
DAFTAR TABEL	xvii
BAB I PENDAHULUAN	I-1
1.1 Latar Belakang	I-1
1.2 Rumusan Masalah	I-4
1.3 Batasan Masalah	I-4
1.4 Tujuan Penelitian	I-4
1.5 Sistematika Penulisan	I-4
BAB II LANDASAN TEORI	II-1
2.1 <i>Intrusion Detection System</i>	II-1
2.2 <i>Knowledge Discovery in Database</i>	II-2
2.3 Klasifikasi	II-5
2.4 <i>Modified K-Nearest Neighbor (MKNN)</i>	II-5
2.4.1 <i>Euclidean Distance</i>	II-6

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.4.2 Validasi <i>Dataset</i>	II-6
2.4.3 Pembobotan <i>Dataset (Weight Voting)</i>	II-7
2.5 Serangan Pada Jaringan Komputer	II-7
2.5.1 <i>Denial of Service</i>	II-8
2.5.2 Port Scanning	II-8
2.5.3 SYN Attack.....	II-9
2.5.4 Teardrop.....	II-9
2.5.5 IP Spoofing	II-10
2.5.6 Smurf Attack.....	II-10
2.6 NSL KDD CUP 1999 <i>Dataset</i>	II-10
2.7 Evaluasi	II-15
2.8 Kajian Penelitian Terkait	II-16
BAB III METODOLOGI PENELITIAN	III-1
3.1 Studi Literatur	III-2
3.2 Perumusan Masalah	III-2
3.3 Pengumpulan Data.....	III-2
3.4 Proses KDD (<i>Knowledge Discovery in Database</i>).....	III-2
3.4.1 <i>Selection</i>	III-2
3.4.2 <i>Pre Processing</i>	III-3
3.4.3 <i>Transformation</i>	III-3
3.4.4 <i>Data Mining</i>	III-3
3.5 Perancangan	III-4
3.6 Implementasi dan Pengujian (<i>Evaluation</i>)	III-4
3.6.1 <i>Black Box</i>	III-4
3.6.2 Pengujian Akurasi	III-4

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3.7 Kesimpulan dan Saran	III-5
BAB IV ANALISA DAN PERANCANGAN	IV-1
4.1 Analisa Data.....	IV-1
4.2 Tahapan KDD (<i>Knowledge Discovery Data</i>)	IV-1
4.2.1 <i>Data Selection</i>	IV-1
4.2.2 <i>Preprocessing data</i>	IV-1
4.2.3 Transformasi Data.....	IV-7
4.2.4 <i>Data Mining</i>	IV-14
4.3 Analisa Sistem	IV-28
4.3.1 <i>Use Case Diagram</i>	IV-28
4.3.2 <i>Activity Diagram</i>	IV-31
4.3.1 <i>Sequence Diagram</i>	IV-36
4.3.2 <i>Class Diagram</i>	IV-39
4.4 Perancangan Sistem	IV-41
4.4.1 Perancangan <i>Database</i>	IV-41
4.4.2 Perancangan Struktur Menu.....	IV-43
4.4.3 Perancangan <i>Interface</i>	IV-43
BAB V IMPLEMENTASI DAN PENGUJIAN	V-1
5.1 Implementasi.....	V-1
5.1.1 Batasan Implementasi	V-1
5.1.2 Lingkungan Implementasi.....	V-1
5.1.3 Implementasi Antarmuka Sistem (<i>Interface</i>).....	V-2
5.2 Pengujian	V-6
5.1.1 Pengujian Akurasi	V-6
5.1.2 Pengujian <i>Black Box</i>	V-10



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

5.1.3 Kesimpulan Pengujian	V-15
----------------------------------	------

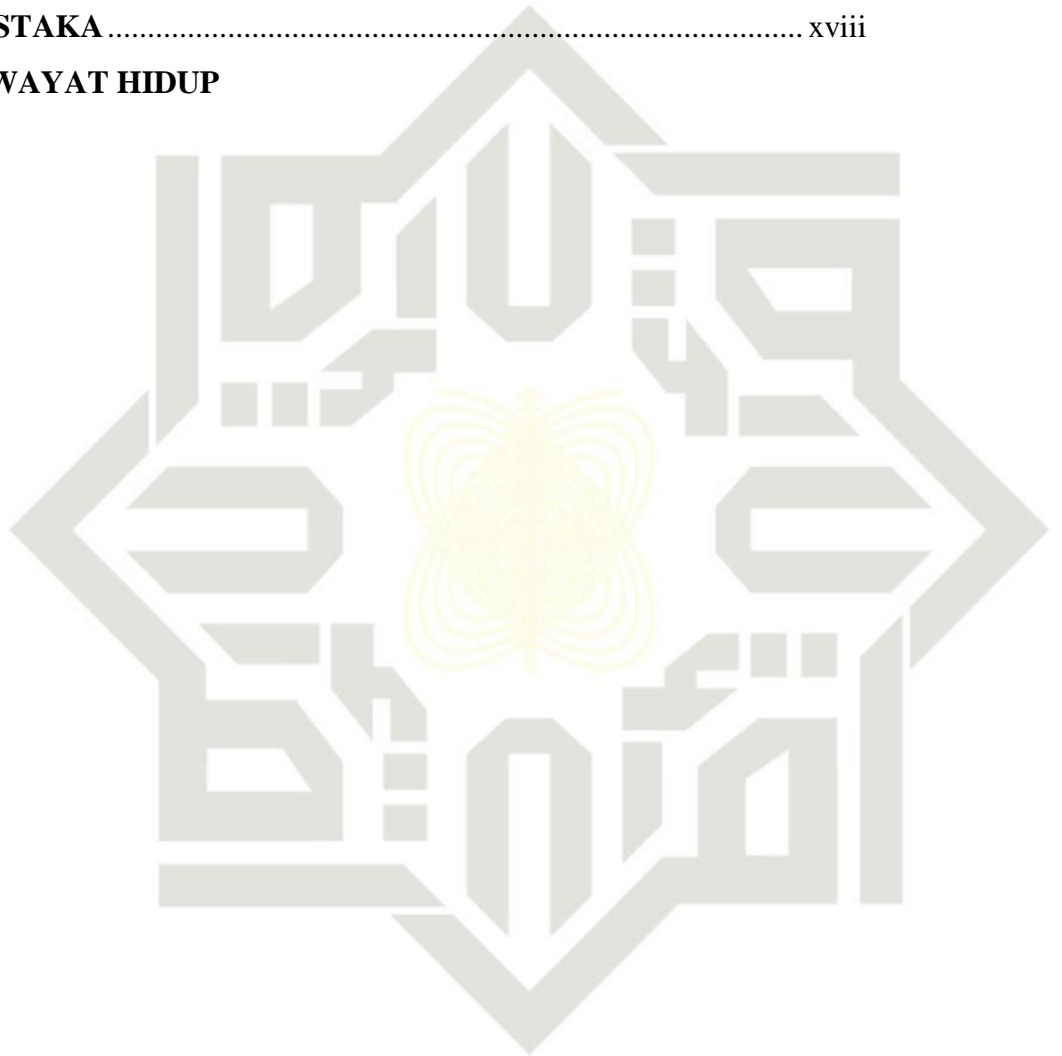
BAB VI PENUTUP	VI-1
-----------------------------	-------------

6.1 Kesimpulan	VI-1
----------------------	------

6.2 Saran	VI-1
-----------------	------

DAFTAR PUSTAKA	xviii
-----------------------------	--------------

DAFTAR RIWAYAT HIDUP	
-----------------------------	--



UIN SUSKA RIAU

DAFTAR GAMBAR

Gambar	Halaman
3.1 Metodologi Penelitian	III-1
3.2 Flowchart MKNN	III-3
4.1 Pencarian dan Penyaringan Nilai Z-score data latih	IV-6
4.2 Penyaringan dan Penyaringan Nilai Z-score data uji	IV-6
4.3 Data Outlier	IV-6
4.4 Data Outlier Csv	IV-7
4.5 Flowchart Klasifikasi MKNN	IV-14
4.6 Use Case Diagram Sistem Klasifikasi Serangan Jaringan	IV-28
4.7 Activity Diagram Import Data Latih	IV-32
4.8 Activity Diagram Import Data Uji	IV-33
4.9 Activity Diagram Pelatihan	IV-34
4.10 Activity Diagram Pengujian	IV-35
4.11 Activity Diagram Pengecekan Hasil	IV-36
4.12 Sequence Diagram Import Data Latih	IV-37
4.13 Sequence Diagram Import Data Uji	IV-37
4.14 Sequence Diagram Pelatihan	IV-38
4.15 Sequence Diagram Pengujian	IV-38
4.16 Sequence Diagram Pengecekan Hasil	IV-39
4.17 Class Diagram Sistem Klasifikasi Serangan	IV-40
4.18 Struktur Menu Sistem	IV-43
4.19 Interface Data Latih	IV-44
4.20 Interface Data Uji	IV-45
4.21 Interface Pelatihan	IV-45
4.22 Interface Pengujian	IV-46
4.23 Interface Hasil	IV-46
4.24 Interface Detail Hasil Pengujian	IV-47

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

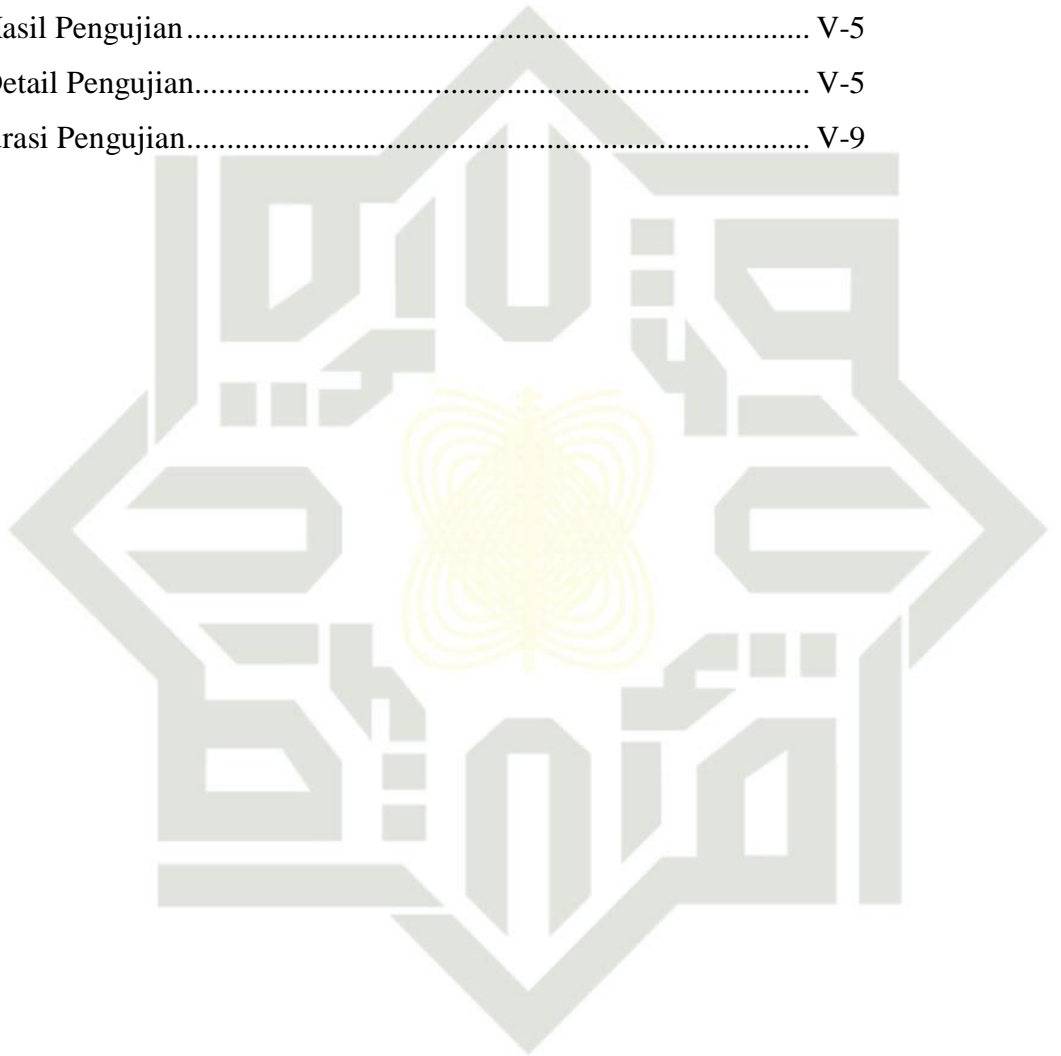
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

5.1	Halaman Data Latih	V-2
5.2	Halaman Data Uji.....	V-3
5.3	Halaman Pelatihan	V-3
5.4	Halaman Pengujian	V-4
5.5	Halaman Hasil Pengujian	V-5
5.6	Halaman Detail Pengujian.....	V-5
5.7	Grafik Akurasi Pengujian.....	V-9



UIN SUSKA RIAU

DAFTAR TABEL

Tabel	Halaman
2.1 Atribut Dataset pada NSL KDD CUP 1999.....	II-11
2.2 Pengelompokan Atribut pada NSL KDD CUP	II-13
2.3 Daftar Kelas Serangan pada NSL KDD CUP	II-15
2.4 Confusion Matrix	II-15
2.5 Penelitian Terkait	II-16
4.1 Data Variabel	IV-2
4.2 Data NSL KDD Variabel 1-41	IV-3
4.3 Transformasi Kelas	IV-7
4.4 Transformasi Protocol, Service dan Flag	IV-7
4.5 Data Transformasi	IV-9
4.6 Hasil Normalisasi	IV-12
4.7 Data Latih NSL-KDD	IV-15
4.8 Data Uji NSL-KDD.....	IV-16
4.9 Jarak Euclidean antar Data Latih	IV-19
4.10 Nilai Validitas Data Latih	IV-23
4.11 Jarak Euclidean Data Uji dan Data Latih.....	IV-24
4.12 Nilai Weight Voting	IV-26
4.13 Mayoritas Data	IV-27
4.14 Hasil Klasifikasi	IV-28
4.15 Use Case Spesifikasi Import data latih.....	IV-29
4.16 Use Case Spesifikasi Import data uji	IV-29
4.17 Use Case Spesifikasi Pelatihan	IV-30
4.18 Use Case Spesifikasi Pengujian	IV-30
4.19 Use Case Spesifikasi Hasil.....	IV-31
4.20 Tabel Data Latih.....	IV-41
4.21 Tabel Data Uji	IV-42
4.22 Tabel Data Hasil.....	IV-42



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

5.1	Hasil Pengujian	V-6
5.2	Confusion Matrix $k=3$	V-9
5.3	Nilai Akurasi Keseluruhan Parameter.....	V-9
5.4	Pengujian Black Box Menu Data Latih	V-10
5.5	Pengujian Black Box Menu Data Uji.....	V-11
5.6	Pengujian Black Box Menu Pelatihan.....	V-12
5.7	Pengujian Black Box Menu Pengujian	V-13
5.8	Pengujian Black Box Menu Hasil Pengujian	V-14

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB I PENDAHULUAN

1.1 Latar Belakang

Penggunaan internet memiliki peranan yang sangat besar dalam mendukung kinerja dan aktivitas manusia dalam kehidupan sehari-hari. Dengan jumlah pengguna pada tahun 2017 berdasarkan data dari APJII (Asosiasi Penyelenggara Jasa Internet Indonesia) sebanyak 143,26 juta jiwa atau sekitar 54,68% dari total populasi penduduk Indonesia yang berjumlah 262 juta (APJII, 2017). Dengan semakin banyaknya jumlah penggunaan internet dalam kehidupan masyarakat, dapat memberikan suatu celah bagi sasaran kejahatan didalam dunia *cyber* yang dilakukan penyusup (Mas'ud Zubairi, 2014). Seperti yang tertera pada lampiran peraturan Menteri Komunikasi dan Informatika Republik Indonesia bahwa pada tahun 2014 total data serangan mencapai 48,4 juta serangan (Renstra Kominfo, 2016). Sehingga dibutuhkan sistem keamanan Intrusion Detection System (IDS) yang mampu memantau dan mendeteksi gangguan atau intrusi pada seluruh sistem.

Intrusion Detection System merupakan perangkat lunak atau sistem yang bekerja secara otomatis untuk memonitor kejadian pada jaringan komputer serta melakukan analisis untuk mengetahui aktivitas pada jaringan komputer tersebut normal atau intrusi (Alder, 2004). IDS sendiri memiliki peranan penting dalam menjaga *integrity* (integritas), *confidentially* (kerahasiaan) dan *availability* (tersediaan) dari sumber jaringan komputer (Alder, 2004). Metode IDS berdasarkan cara untuk mendeteksi serangan ialah berbasis aturan (*Signature Based Detection*) atau berbasis anomali (*Anomaly Based Detection*) dan *statateful protocol analysis* (Liao dkk, 2013).

IDS mendeteksi serangan dengan menganalisis informasi dari sebuah *host* tunggal pada beberapa lokasi dalam lingkup seluruh jaringan (Mas'ud Zubairi, 2014). Sehingga komponen IDS dapat melewati komunikasi antara satu sama lain yang mengakibatkan terhambatnya kemampuan untuk mendeteksi serangan



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

terdistribusi dalam skala besar (Mas'ud Zubairi, 2014). Hal ini menyebabkan menurunnya fungsi IDS yaitu mengklasifikasikan aktifitas jaringan apakah termasuk kategori normal atau intrusi (Mas'ud Zubairi, 2014). Penelitian (Liao dkk, 2013) dalam survei reviewnya tentang IDS menyatakan perlu adanya pengetahuan dan terobosan baru untuk memperbaiki keterbatasan teknik, performa dan kinerja IDS. *Data mining* dapat digunakan sebagai langkah alternatif untuk mengatasi permasalahan pada IDS (Mongkareng, dkk 2017). Banyak masalah baru yang muncul telah diselesaikan dengan metode *data mining* seperti masalah statistik, algoritma komputasi, pengenalan pola serta *machine learning* (Yang dan Wu, 2006). Terdapat beberapa algoritma *data mining* yang dapat digunakan, salah satunya ialah menggunakan algoritma *K-Nearest Neighbor* (KNN).

KNN merupakan teknik klasifikasi yang melakukan prediksi secara tegas pada data uji berdasarkan perbandingan k tetangga terdekat (Prasetyo, 2012). Selain itu KNN merupakan algoritma yang menggunakan seluruh data latih untuk melakukan proses klasifikasi (Prasetyo, 2012). KNN memiliki beberapa kelebihan diantaranya ialah pelatihan sangat cepat, sederhana dan mudah dipelajari, tahan terhadap data pelatihan yang memiliki derau serta efektif jika dapat pelatihan besar (Wirawan dan Eksistyanto, 2015). Namun, KNN juga memiliki beberapa kelemahan, yakni komputasi yang cukup tinggi, dikarenakan harus menghitung jarak di setiap *query* untuk semua sampel pelatihan, membutuhkan memori yang besar, tingkat akurasi yang rendah ketika data pelatihan memiliki multi dimensi dan jarak yang berdasarkan pembelajaran tidak jelas (Parvin dkk, 2010). Penelitian terkait untuk perbaikan KNN telah banyak ditemui, baik dalam memperbaiki nilai akurasi KNN maupun dalam hal optimasi nilai k pada KNN (Mutrofin dkk, 2014). Salah satunya ialah metode Modified K-Nearest Neighbor (MKNN).

Algoritma *Modified K-Nearest Neighbor* (MKNN) ialah pengembangan performansi dari metode klasifikasi sebelumnya yakni *K-Nearest Neighbor* (KNN) (Parvin dkk, 2008). Konsep utama dari metode ini ialah pengklasifikasian dari *sample* uji berdasarkan tag tetangganya / *neighbor*. Pada MKNN terdiri atas 2 pemrosesan, pertama validasi data *training* dan yang kedua mengimplementasikan pembobotan KNN (Parvin dkk, 2008). Nilai k sangat mempengaruhi hasil



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

keakurasian data serta harus melalui serangkaian proses percobaan terlebih dahulu. Tingkat akurasi MKNN terbukti lebih tinggi jika dibandingkan terhadap metode klasifikasi pendahulunya yakni KNN (Parvin dkk, 2010). Hal ini dibuktikan dari penelitian sebelumnya pada *dataset Balance Scale*, metode KNN memiliki tingkat akurasi sebesar 83,22% sedangkan metode MKNN memiliki tingkat akurasi sebesar 87,10% (Parvin dkk, 2010). Selanjutnya penelitian yang dilakukan oleh (Okfalisa dkk, 2017) pada *dataset* transfer tunai bersyarat program keluarga harapan yang terdiri dari 7395 *records*. Pada metode KNN akurasi tertinggi diperoleh sebesar 94,95% dengan rata-rata akurasi selama proses pengujian ialah 93,94% sedangkan pada metode MKNN akurasi tertinggi diperoleh sebesar 99,51% dengan rata-rata akurasi sebesar 99.20%.

Penelitian tentang klasifikasi *Modified K-Nearest Neighbor* telah dilakukan sebelumnya. Penelitian yang dilakukan oleh (Wafiyah dkk, 2017). Penelitian ini membahas tentang penerapan metode MKNN untuk mengklasifikasikan gejala penyakit demam. Nilai rata-rata akurasi pengujian pengaruh nilai k terhadap akurasi ialah 88,55% dengan akurasi tertinggi pada pengujian nilai $k=3$ dengan akurasi sebesar 94,95%.

Penelitian selanjutnya ialah penelitian yang dilakukan oleh (Fernanda dkk, 2017). Penelitian ini membahas tentang penerapan metode MKNN untuk mengklasifikasikan gejala penyakit diabetes mellitus. Hasil pada penelitian ini diperoleh akurasi terbaik sebesar 93,33% dengan menggunakan nilai $k=3$.

Penelitian berikutnya ialah penelitian yang dilakukan oleh (Simanjutak dkk, 2017). Penelitian ini membahas tentang penerapan metode MKNN untuk mengklasifikasikan jenis penyakit yang menyerang tanaman kedelai. Hasil akurasi tertinggi pada penelitian ini diperoleh sebesar 100% dengan nilai $k=1$ dan rata-rata akurasi dari 5 percobaan sebesar 98,83%.

Pada penelitian ini akan menggunakan *Dataset* NSL-KDD Cup 1999 (<http://www.unb.ca/cic/datasets/ns1.html>). NSL KDD CUP merupakan *dataset* yang diusulkan untuk mengurangi masalah yang mendasar pada KDD CUP 1999. Kekurangan yang mendasar pada *dataset* KDD CUP 1999 yakni banyaknya *record* yang redundan. Ditemukan sekitar 78% dan 75% *record* yang duplikat pada KDD

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Train dan *test set* (Tavallae dkk, 2009). Terdapat beberapa keuntungan menggunakan *dataset* ini yaitu : tidak ada *record* yang berlebihan di dalam *train set*, sehingga *classifier* tidak akan menghasilkan hasil yang bias serta tidak ada duplikat *record* pada *test set* yang memiliki *reduction rates* yang lebih baik (Wirawan dan Eksistyanto, 2015).

Berdasarkan penjabaran permasalahan yang telah dipaparkan, maka akan dilakukan penerapan metode MKNN untuk mengklasifikasikan data serangan jaringan komputer pada *dataset* NSL KDD CUP 1999. Diharapkan pada penelitian ini dihasilkan akurasi deteksi serangan jaringan komputer yang optimal.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, rumusan masalah pada penelitian ini bagaimana penerapan metode *Modifiked K-Nearest Neighbor* untuk mengklasifikasikan data serangan jaringan komputer NSL-KDD.

1.3 Batasan Masalah

Dalam penelitian penerapan metode MKNN ini dibatasi pada:

1. Data yang digunakan berasal dari *dataset* NSL-KDD Cup 1999 dengan data latih berjumlah 125.974 dan data uji berjumlah 22.543
 2. Kelas yang digunakan ialah normal, serangan *probe*, serangan *Denial of Service* (DOS), serangan *remote to local* (R2L) dan serangan *user to root* (U2R).
- Parameter yang digunakan berjumlah 29 parameter.

1.4 Tujuan Penelitian

Tujuan penelitian ini ialah untuk menghasilkan aplikasi yang menerapkan metode MKNN untuk klasifikasi serangan jaringan serta menguji tingkat akurasi dari metode yang digunakan tersebut.

1.5 Sistematika Penulisan

Berikut merupakan susunan sistematika penulisan laporan tugas akhir yang akan disusun sebagai berikut :



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB I

PENDAHULUAN

Berisi dasar-dasar dari penulisan atau deskripsi umum pada tugas akhir yang meliputi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian dan sistematika penulisan

BAB II

LANDASAN TEORI

Berisi penjelasan mengenai dasar-dasar teori, rujukan dan metode yang digunakan sebagai dasar dalam menerapkan metode

BAB III

METODOLOGI PENELITIAN

Berisi penjelasan mengenai tahapan dalam tugas akhir yang dilakukan penulis, dalam menerapkan metode

BAB IV

ANALISA DAN PERANCANGAN

Berisi tentang mengenai pengolahan *dataset* NSL-KDD Cup 1999 hingga analisa metode metode MKNN dalam memproses data. Sehingga menghasilkan solusi yang diinginkan secara manual serta perancangan sistem *database* serta interface yang akan dibangun.

BAB V

IMPLEMENTASI DAN PENGUJIAN

Berisi tentang tahap pembangunan dari aplikasi berdasarkan pada rancangan yang dijelaskan pada bab sebelumnya. Kemudian akan diuji baik fungsi maupun akurasi menggunakan *blackbox* dan *confusion matrix*.

BAB VI

PENUTUP

Berisi tentang kesimpulan serta saran terhadap hasil penelitian yang telah dilakukan.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB II LANDASAN TEORI

2.1 *Intrusion Detection System*

Intrusion Detection System merupakan perangkat lunak atau sistem yang bekerja secara otomatis untuk memonitor kejadian pada jaringan komputer serta melakukan analisis untuk mengetahui aktivitas pada jaringan komputer tersebut normal atau intrusi (Alder, 2004).

Suatu IDS dapat didefinisikan sebagai *tool*, metode atau sumber daya yang memberikan bantuan untuk melakukan identifikasi, memberikan laporan terhadap aktivitas jaringan komputer. Pada dasarnya IDS tidak mendeteksi penyusup tetapi hanya mendeteksi aktivitas *traffic* jaringan yang tidak layak terhadai sehingga awal dari langkah kerja penyerang bisa diketahui. Dengan demikian *network administrator* dapat melakukan tindakan pencegahan dan bersipa atas kemungkinan serangan yang akan terjadi (Alder, 2004).

Ada 2 jenis IDS yaitu *Host Intrusion Detection System* (HIDS) dan *Network Based Intrusion Detection System* (NIDS).

1. *Host Intrusion Detection System* (HIDS)

HIDS ditempatkan sebagai agen dari sebuah host. *Intrusion Detection System* jenis ini dapat melihat ke dalam *file log* dari sistem dan aplikasi untuk mendeteksi segala kegiatan intrusi. Beberapa dari sistem ini bersifat reaktif, yang berarti HIDS hanya melaporkan *user* ketika serangan telah terjadi. HIDS bekerja pada *host* yang akan dilindungi. Pada IDS jenis ini dapat melakukan berbagai macam tugas untuk mendeteksi adanya serangan yang dilakukan pada *host* tertentu. Keunggulan yang terdapat pada HIDS ialah tugas-tugas yang berhubungan dengan keamanan *file*. Misalnya ada atau tidaknya *file* yang telah dirubah atau ada usaha untuk mendapatkan akses kepada *file* yang bersifat sensitif (Mydza, 2011).



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta dilindungi undang-undang UIN Suska Riau
Sultan Syarif Kasim Riau

2.1 Network Intrusion Detection System (NIDS)

NIDS adalah *Intrusion Detection System* yang menangkap paket data yang bergerak di suatu media jaringan (kabel, nirkabel) dan mencocokkan paket data tersebut dengan *signature* yang terletak pada *database*. Bila paket data cocok dengan *signature* pada *intruder* maka peringatan akan dibuat dan paket data akan disimpan ke sebuah file atau database. NIDS digunakan untuk melakukan monitoring di seluruh segmen jaringan NIDS akan mengumpulkan paket-paket data yang terdapat pada jaringan kemudian mengalisanya serta menentukan apakah paket-paket tersebut berupa suatu paket yang normal atau suatu aktivitas yang mencurigakan (Mydza, 2011).

Intrusion Detection System terbagi ke dalam dua kategori pendekatan : *intrusion detection system* berbasis *signature* dan *anomaly detection system* berbasis profil dan pengenalan aturan dari suatu sistem. *Pattern* atau *rule* adalah bagian paling penting dari *Intrusion Detection System*.

2.2 Knowledge Discovery in Database

Knowledge Discovery in Database (KDD) adalah proses menentukan informasi yang berguna serta pola-pola yang ada dalam data. Informasi ini terkandung dalam basis data yang berukuran besar yang sebenarnya tidak diketahui dan potensi bermanfaat (Han dkk, 2012). Tahapan proses KDD terdiri dari:

1. Data Selection

Pada proses ini dilakukan pemilihan himpunan data, menciptakan himpunan data target, atau memfokuskan pada *subset* variabel, (sampel data) dimana penemuan akan dilakukan (Han dkk, 2012). Hasil seleksi disimpan dalam suatu berkas yang terpisah dari basis data operasional.

2. Pre-Processing

Pre-Processing data dilakukan dengan membuang data yang tidak konsisten dan *noise*, duplikasi data, memperbaiki kesalahan data dan data yang bias (Han dkk, 2012) . Berikut beberapa tahapan *preprocessing*:

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

a. Data Cleaning

Proses ini merupakan pembersihan data yang memiliki *noise*. Dengan pembersihan ini dapat dilihat data mana yang memiliki kesalahan sehingga dapat diperbaiki sebelum menjadi inputan data pada proses data mining berikutnya (Han dkk, 2012). Berikut beberapa cara menanggulangi hal tersebut (Han dkk, 2012):

1. Mengabaikan tupel dari kelas yang hilang
2. Mengisi data yang hilang secara manual
3. Menggantikan konstanta yang umum yaitu mengganti semua nilai atribut yang hilang dengan konstanta yang sama seperti label.
4. Menggunakan ukuran tendency tengah atribut untuk mengisi *missing value* seperti mean atau median.
5. Menggunakan kemungkinan nilai untuk mengisi *missing value* dengan regresi.

b. Data Integration

Data Integration merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru. Integrasi data ini dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik (Han dkk, 2012).

3. Transformation

Proses ini mentransformasikan atau menggabungkan data ke dalam yang lebih tepat untuk melakukan proses mining dengan cara melakukan peringkasan. Terdapat beberapa cara yang dapat dilakukan dalam transformasi data salah satunya normalisasi (Han dkk, 2012). Pada proses normalisasi nilai pada data atribut akan dibuat dalam skala 0 hingga 1 sehingga sebaran data tidak terlalu jauh. Proses normalisasi dilakukan dengan menggunakan persamaan *min-max normalization*, dimana x' adalah hasil nilai normalisasi, $\min(x)$ adalah nilai minimum pada atribut x dan $\max(x)$ ialah nilai maksimum pada atribut x (Larose, 2005). Berikut persamaan proses normalisasi dengan menggunakan persamaan *min-max normalization* (Larose, 2005):

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \dots \dots \dots (2.1)$$

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Keterangan :

x_i = nilai setelah dinormalisasi

x = nilai sebelum dinormalisasi

$\min(x)$ = nilai minimum atribut

$\max(x)$ = nilai maksimum atribut

Proses normalisasi juga dilakukan menggunakan persamaan *Z-score*, pada persamaan ini melibatkan nilai mean serta standar deviasi terhadap data *input* (Larose, 2005). Berikut ialah persamaan *Z-score* (Larose, 2005):

$$x_i = \frac{x - \text{mean}(x)}{SD(x)} \dots \dots \dots (2.2)$$

Keterangan :

x = nilai data input

$\text{mean}(x)$ = rata-rata dari tiap kolom data input

$SD(x)$ = standar deviasi dari tiap kolom data input

Berikut ialah persamaan yang digunakan untuk mencari nilai standar deviasi (Soewarno, 1995):

$$SD(x) = \sqrt{\frac{\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}{n(n-1)}} \dots \dots \dots (2.3)$$

Keterangan :

x_i = nilai x ke - i

n = ukuran data

4 Data Mining

Proses *data mining* yaitu proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik, metode atau algoritma tertentu sesuai dengan tujuan dari proses KDD secara keseluruhan (Han dkk, 2012).²

4.1 Interpretational Evaluasi

Proses untuk menerjemahkan pola-pola yang dihasilkan dari data mining. Mengevaluasi (menguji) apakah pola atau informasi yang ditemukan bersesuaian atau bertentangan dengan fakta atau hipotesa sebelumnya. Pengetahuan yang diperoleh dari pola-pola yang terbentuk dipresentasikan dalam bentuk visualisasi (Han dkk, 2012).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

6 Knowledge

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Dalam proses ini biasanya melibatkan orang-orang yang tidak memahami *data mining* (Han dkk, 2012).

2.3 Klasifikasi

Klasifikasi adalah proses penemuan model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui (Han dkk, 2012). Klasifikasi data terdiri dari 2 proses. Pertama ialah *learning (face training)*, dimana algoritma klasifikasi dibuat untuk menganalisa data *training* lalu dipresentasikan dalam bentuk *rule* klasifikasi, dimana data tes digunakan untuk memperkirakan akurasi dari *rule* (Han dkk, 2012).

2.4 Modified K-Nearest Neighbor (MKNN)

Algoritma Modified K-Nearest Neighbor (MKNN) ialah pengembangan dari metode KNN yang diusulkan oleh (Parvin dkk, 2008) yang sebagian bertujuan untuk mengatasi masalah tingkat akurasi yang rendah pada algoritma KNN. Pengembangan dilakukan dengan melakukan modifikasi pada algoritma KNN yang bertujuan untuk meningkatkan kinerja algoritma KNN. Ide utama dari pengembangan algoritma KNN yang dilakukan adalah untuk menggunakan tetangga yang kuat dalam *dataset*. Pada metode MKNN menambahkan proses validasi pada setiap *dataset* setelah dilakukan nya perhitungan jarak antara titik pada data *training* (x) dan titik pada data *testing* (y) dengan menggunakan *euclidean distance*. Selanjutnya proses klasifikasi dijalankan dengan melakukan pembobotan pada *dataset* dengan menggunakan nilai validasi sebagai faktor perkalian (Parvin dkk, 2008).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.4.1 Euclidean Distance

Berikut merupakan rumus *Euclidean* yang digunakan untuk mendefinisikan jarak antara titik pada data *training* (x) dan titik pada data *testing* (y) (Wilson dan Martinez, 1997):

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots \dots \dots (2.4)$$

Keterangan:

- d(x,y) = jarak Euclidean
- x_i = data ke-i
- y_i = data ke-j
- i = 1,2,3,...,n

2.4.2 Validasi Dataset

Pada algoritma *Modified K-Nearest Neighbor dataset* akan melalui tahap validasi terlebih dahulu/ validasi setiap titik dihitung sesuai dengan tetangganya dan dilakukan hanya satu kali (Parvin dkk, 2008). Nilai validasi setiap *dataset* kemudian akan digunakan untuk melakukan pembobotan pada *dataset* yang akan digunakan untuk menentukan kelas suatu data *testing*. Untuk memvalidasi sebuah *dataset*, harus ditentukan terlebih dahulu parameter H yang akan digunakan. Parameter H ini merupakan sebuah nilai yang mempresentasikan jumlah *neighbor* yang digunakan untuk melakukan proses validasi. Diantara H *neighbor* terdekat dari *dataset* x, nilai validasi (x) setiap *dataset* dilakukan dengan menghitung jumlah data yang memiliki label yang sama dengan label x. Adapun rumus untuk menghitung nilai validasi setiap *dataset* ialah (Parvin dkk, 2010):

$$validity(x) = \frac{1}{H} \sum_{i=1}^H S(lbl(x), lbl(N_i(x))) \dots \dots \dots (2.5)$$

Keterangan :

- $N_i(x)$: nilai dari tetangga yang digunakan
- lbl(x) : label dari *dataset*
- $l_i(x)$: label dari tetangga terdekat dengan i = 1,2,3.... H.
- S : Fungsi yang memperhitungkan kesamaan antara satu *dataset* dan *dataset* lainnya. Berikut ialah definisi dari fungsi S:

$$S(a.b) = \begin{cases} 1 & a=b \\ 0 & a \neq 0 \end{cases} \dots \dots \dots (2.6)$$

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Keterangan :

a = kelas a pada data latih

b = kelas lain selain a pada data latih

2.4.3 Pembobotan *Dataset* (*Weight Voting*)

Pembobotan *dataset* dalam KNN merupakan salah satu dari variasi metode KNN. Variasi metode KNN ini melakukan penentuan kelas dari data objek baru dengan tidak melakukan *voting* mayoritas kelas data pada K tetangga terdekat, melainkan dengan cara melakukan *voting* berdasarkan bobot pada *dataset*. setiap *dataset* akan diberikan bobot dengan melakukan perhitungan (Parvin dkk, 2010). Misalkan, bobot setiap *dataset* dihitung dengan menggunakan persamaan:

$$w = \frac{1}{de+1} \dots \dots \dots (2.7)$$

w merupakan bobot dan d_e merupakan jarak *Euclidean*. Bobot ini kemudian akan dijumlahkan untuk setiap kelas dan kelas dengan bobot tertinggi akan dipilih untuk menentukan kelas dari data objek yang baru.

Pada metode *Modified K-Nearest Neighbor* bobot masing-masing tetangga dilakukan dengan perhitungan sebagai berikut :

$$w = \frac{1}{de+a} \dots \dots \dots (2.8)$$

a merupakan sebuah konstanta yang disebut sebagai *smoothing regulator*. Kemudian nilai yang diperoleh dari perhitungan diatas akan dikalikan dengan nilai validasi untuk mendapatkan nilai bobot akhir (Parvin dkk, 2010). Berikut ialah perhitungan bobot akhir :

$$w(i) = validity(i) \times \frac{1}{de+a} \dots \dots \dots (2.9)$$

$w(i)$ dan $validity(i)$ merupakan bobot dan nilai validasi dari I tetangga terdekat antara *dataset* dan data testing. Teknik pembobotan diatas akan memberikan nilai yang tinggi untuk *dataset* yang memiliki nilai validasi besar serta kesamaan yang tinggi dengan data *testing* (Parvin dkk, 2010).

2.5 Serangan Pada Jaringan Komputer

Berikut serangan-serangan yang terjadi pada jaringan komputer diantaranya ialah sebagai berikut:



2.5.1 Denial of Service

Merupakan jenis serangan yang dilakukan untuk membuat komputer atau jaringan komputer tidak dapat menyediakan layanan secara normal (Ariyus, 2006). Umumnya DoS menargetkan serangan pada *bandwidth* jaringan komputer atau koneksi jaringan (*connectivity*). *Bandwidth attack* membanjiri jaringan dengan volume *traffic* yang tinggi, sehingga semua *resources* (sumber daya) yang ada, tidak dapat melayani *request* (permintaan) dari *legitimate user* (user yang salah) (Ariyus, 2006). Serangan DoS ditandai oleh adanya usaha *attacker* untuk mencegah *legitimate user* dari penggunaan *resource* yang diinginkan. Berikut ialah cara penyerangan yang dilakukan pada DoS *attack* (Ariyus, 2006):

- a. Mencoba untuk membanjiri (*flood*) *network*, dengan demikian mencegah lalu lintas yang *legitimate* pada jaringan.
Mencoba mengganggu koneksi antar dua mesin, dengan demikian mencegah suatu akses layanan.
- b. Mencoba untuk mencegah individu tertentu dari mengakses layanan. Mencoba untuk mengganggu layanan sistem yang spesifik.

2.5.2 Port Scanning

Port scanning merupakan suatu proses untuk mencari dan membuka *port* pada suatu jaringan komputer, dari hasil *scanning* akan di dapat kelemahan dari sistem tersebut. Pada dasarnya sistem *port scanning* mudah untuk dideteksi, namun penyerang menggunakan berbagai metode untuk menyembunyikan serangan dengan menggunakan *port scanning* (Ariyus, 2006). Sebagai contoh banyak dari jaringan tidak membuat *file log connection*, sehingga penyerang dapat mengirimkan *initial packet* dengan suatu SYN tapi tidak ada ACK, dan mendapatkan respon kembali (Selain SYN jika suatu port terbuka) dan kemudian berhenti pada port tersebut. Hal ini sering disebut dengan SYN *scan* atau *half open scan* (Ariyus, 2006).

Penyerang akan mengirim paket lain pada *port* yang masih belum mendapatkan RST dan akan membuka *port* tersebut namun tidak terjadi respon apapun pada *file log* atau kesalahan dari *file* dan *device* lainnya (Ariyus, 2006).

2.5.3 SYN Attack

Serangan yang dilakukan apabila ditemukan kelemahan dari spesifikasi TCP/IP, paket SYN dikirimkan pada saat memulai *handshake* antara aplikasi sebelum pengiriman data dilakukan. Proses terjadi ketika sebuah koneksi TCP dimulai, *host* tujuan akan menerima paket SYN (*Synchronize*) dari *host* pengirim dan akan mengirim kembali paket SYN ACK (*synchronize acknowledge*) (Ariyus, 2006). *Host* tujuan kemudian akan mendapatkan paket ACK sebagai kelanjutan SYN ACK sebelum koneksi tersebut terhubung secara lengkap. Pada saat menunggu paket ACK, akan terjadi sebuah proses antrian koneksi pada *host* tujuan yang berada dalam kondisi menunggu koneksi. Antrian ini biasanya akan cepat hilang karena paket ACK akan datang beberapa *millisecond* setelah SYN ACK (Ariyus, 2006).

Serangan TCP SYN memanfaatkan kelemahan desain koneksi TCP dengan membuat *host* pengirim secara acak ke *host* tujuan dan *host* tujuan akan mengirim kembali paket SYN ACK ke alamat pengirim dan menambahkan pada antrian terhadap alamat yang sebenarnya tidak ada, sehingga jika hal ini terjadi berulang-ulang maka akan mengakibatkan paket yang tidak mempunyai alamat yang sebenarnya berada dalam banyak antrian TCP dan membuat layanan TCP seperti FTP maupun WWW sibuk dan akhirnya berhenti (Ariyus, 2006).

2.5.4 Teardrop

Merupakan suatu teknik yang dikembangkan dengan cara mengeksploitasi proses *disassembly-reassembly* paket data (Ariyus, 2006). Dalam jaringan internet, data harus dipotong menjadi kecil-kecil guna untuk menjamin reliabilitas dan proses *multiple* akses jaringan. Pada proses pemotongan paket data yang normal setiap potongan diberikan informasi *offset* data yang berbunyi “potongan paket ini merupakan potongan 600 byte dari total 800 byte paket yang dikirim” (Ariyus, 2006). Program *Teardrop* akan memanipulasi *offset* potongan data sehingga terjadi *overlapping* antara paket yang diterima pada bagian penerima setelah potongan paket-paket ini di *reassembly*. *Overlapping* dapat menimbulkan sistem *crash*, *hang* dan *reboot* (Ariyus, 2006).



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.5.5 IP Spoofing

IP *Spoofing* adalah suatu serangan teknis yang rumit terdiri dari beberapa komponen (Ariyus, 2006). Serangan ini ialah eksploitasi keamanan yang bekerja dengan menipu komputer. Seolah-olah yang menggunakan komputer tersebut ialah orang lain. Hal ini terjadi karena *design flaw* lubang keamanan yang dapat dikategorikan ke dalam kesalahan desain ialah desain urutan nomor dari paket TCP/IP. Kesalahan ini dapat dieksploitasi sehingga menimbulkan masalah (Ariyus, 2006).

2.5.6 Smurf Attack

Serangan jenis ini biasanya dilakukan dengan menggunakan IP *Spoofing*, yaitu mengubah alamat IP dari datannya *request* (Ariyus, 2006). Selain itu serangan ini merupakan serangan secara paksa pada fitur spesifikasi alamat IP yang dikenal sebagai *direct broadcast addressing*. Seorang *Smurf hacker* biasanya membanjiri router dengan paket permintaan *echo Internet Control Message Protocol* (ICMP) dengan menggunakan utilitas ping (Ariyus, 2006). Karena alamat IP tujuan pada paket yang dikirim adalah alamat *broadcast* dari jaringan, maka router akan mengirimkan permintaan ICMP *echo* ini ke semua mesin yang ada pada jaringan. Jika terdapat banyak *host* di jaringan, maka akan terjadi trafik ICMP *echo response* dan permintaan dalam jumlah yang sangat besar (Ariyus, 2006).

2.6 NSL KDD CUP 1999 Dataset

NSL KDD CUP merupakan *dataset* yang diusulkan untuk mengurangi masalah yang mendasar pada KDD CUP 1999. Kekurangan yang mendasar pada *dataset* KDD CUP 1999 yakni banyaknya *record* yang redundan. Ditemukan sekitar 78% dan 75% *record* yang duplikat pada KDD *Train* dan *test set* (Tavallaei et al., 2009). Keuntungan menggunakan NSL KDD CUP 1999 ialah (Dhanabal dan Shanatharajah, 2015):

- a. Tidak ada *record* yang berlebihan pada *train set*, sehingga *classifier* tidak akan menghasilkan hasil yang bias
- b. Tidak ada duplikat *record* pada *test set* yang memiliki *reduction rates* lebih baik

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

c. Jumlah *record* yang dipilih dari setiap level grup yang berbeda berbanding terbalik dengan persentase *record* pada *dataset* KDD asli

NSL KDD CUP 1999 terdiri dari 41 atribut (Tabel 2.1), 1 kelas normal dan 39 jenis serangan yang dikelompokkan menjadi 4 kategori yaitu DoS, R2L, U2R dan Probe (tabel 2.3).

Tabel 2.1 Atribut Dataset pada NSL KDD CUP 1999

No	Atribut	Deskripsi	Data Sampel
1	<i>Duration</i>	Lama waktu Koneksi	0
2	<i>Protocol_type</i>	Tipe protocol yang digunakan pada koneksi	TCP
3	<i>Service</i>	Tujuan <i>Network Service</i> yang digunakan	ftp_data
4	<i>Flag</i>	Status koneksi (normal atau error)	SF
5	<i>Src_bytes</i>	Jumlah <i>bytes</i> dari sumber ke tujuan	491
6	<i>Dst_bytes</i>	Jumlah <i>bytes</i> dari tujuan ke sumber	0
7	<i>Land</i>	Jika sumber dan tujuan alamat IP dan port numbers sama maka variabel ini akan bernilai 1 <i>else</i> 0	0
8	<i>Wrong_fragment</i>	Jumlah <i>Wrong_Fragment</i> pada koneksi	0
9	<i>Urgent</i>	Jumlah paket yang urgen pada koneksi	0
10	<i>Hot</i>	Jumlah indikator <i>hot</i> pada <i>content</i> (mengakses sistem direktori, membuat <i>programs</i> dan menjalankan <i>programs</i>)	0
11	<i>Num_failed_logins</i>	Jumlah Percobaan login yang gagal	0
12	<i>Logged_in</i>	Status login (bernilai 1 jika login sukses dan bernilai 0 jika gagal)	0
13	<i>Num_compromised</i>	Jumlah kondisi <i>compromised</i>	0
14	<i>Root_shell</i>	Bernilai 1 jika <i>root shell</i> didapat dan bernilai 0 jika tidak	0
15	<i>Su_attempted</i>	Bernilai 1 jika dilakukan percobaan atau menjalankan perintah “ <i>su root</i> ” dan bernilai 0 apabila tidak dijalankan	0
16	<i>Num_roott</i>	Jumlah <i>root</i> yang diakses atau jumlah operasi yang diakses sebagai <i>root</i> pada koneksi	0
17	<i>Num_file_creations</i>	Jumlah operasi <i>file creations</i> pada koneksi	0
18	<i>Num_shells</i>	Jumlah <i>shell prompts</i>	0

19	<i>Num_access_files</i>	Jumlah operasi pada akses file control	0
20	<i>Num_outbound_cmds</i>	Jumlah <i>outbound command</i> pada <i>ftp session</i>	0
21	<i>Is_hot_login</i>	Bernilai 1 jika login yang terdapat pada list <i>hot</i> (root or admin)	0
22	<i>Is_guest_login</i>	Bernilai 1 jika login sebagai “ <i>guest login</i> ” dan sebaliknya	0
23	<i>Count</i>	Jumlah koneksi yang mengarah ke tujuan host yang sama pada jaringan yang sedang berjalan.	2
24	<i>Srv_count</i>	Jumlah koneksi dengan <i>service</i> yang sama (<i>port number</i>) pada koneksi yang sedang berjalan.	2
25	<i>Serror_rate</i>	Persentase koneksi yang diaktifkan <i>flag</i> (4) s0, s1, s2 atau s3 diantara <i>connection aggregated</i> pada <i>count</i> (23)	0
26	<i>Srv_serror_rate</i>	Persentase koneksi yang diaktifkan <i>flag</i> (4) so, s1, s2 atau s3 diantara <i>connection aggregated</i> pada <i>srv_count</i> (24)	0
27	<i>Rerror_rate</i>	Persentase koneksi yang diaktifkan <i>flag</i> (4) REJ, diantara <i>connection aggregated</i> pada <i>count</i> (23)	0
28	<i>Srv_rerror_rate</i>	Persentase koneksi yang diaktifkan <i>flag</i> (4) REJ, diantara <i>connection aggregated</i> pada <i>srv_count</i> (24)	0
29	<i>Same_srv_rate</i>	Persentase koneksi pada <i>service</i> yang sama, diantara <i>connection aggregated</i> pada <i>count</i> (23)	1
30	<i>Diff_srv_rate</i>	Persentase koneksi pada <i>service</i> yang berbeda, diantara <i>connection aggregated</i> pada <i>count</i> (23)	0
31	<i>Srv_diff_host_rate</i>	Persentase koneksi ke tujuan <i>machines</i> yang berbeda diantara <i>connection aggregated</i> pada <i>srv_count</i> (24)	0
32	<i>Dst_host_count</i>	Jumlah koneksi yang memiliki <i>host ip address</i> tujuan yang sama	150
33	<i>Dst_host_srv_count</i>	Jumlah koneksi yang memiliki <i>port number</i> yang sama	25
34	<i>Dst_host_same_srv_rate</i>	Persentase koneksi dengan <i>services</i> yang sama diantara	0.17

Hak Cipta Dilindungi Undang-Undang

- Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
- Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

		<i>connections aggregated</i> pada <i>dst_host_count</i> (32)	
35	<i>Dst_host_diff_srv_rate</i>	Persentase koneksi dengan <i>services</i> yang berbeda, diantara <i>connections aggregated</i> pada <i>dst_host_count</i> (32)	0.03
36	<i>Dst_host_same_src_port_rate</i>	Persentase koneksi dengan sumber <i>port</i> yang sama diantara <i>connections aggregated</i> pada <i>dst_host_srv_count</i> (33)	0.17
37	<i>Dst_host_srv_diff_host_rate</i>	Persentase koneksi dengan tujuan <i>machines</i> yang berbeda, diantara <i>connections aggregated</i> pada <i>dst_host_srv_count</i> (33)	0
38	<i>Dst_host_serror_rate</i>	Persentase koneksi yang diaktifkan <i>flag</i> (4) s0, s1, s2 atau s3 diantara <i>connections aggregated</i> pada <i>dst_host_count</i> (32)	0
39	<i>Dst_host_srv_serror_rate</i>	Persentase koneksi yang diaktifkan <i>flag</i> (4) s0, s1 atau s3 diantara <i>connections aggregated</i> pada <i>dst_host_srv_count</i> (33)	0
40	<i>Dst_host_rerror_rate</i>	Persentase koneksi yang diaktifkan <i>flag</i> (4) REJ, diantara <i>connections aggregated</i> pada <i>dst_host_count</i> (32)	0.05
41	<i>Dst_host_srv_rerror_rate</i>	Persentase koneksi yang diaktifkan <i>flag</i> (4) REJ, diantara <i>connections aggregated</i> pada <i>dst_host_srv_count</i> (33)	0

Sumber : (Dhanabal dan Shantharajah, 2015)

Berikut pembagian 41 atribut diatas yang dikelompokkan menjadi 3 tipe data, yang dapat dilihat pada Tabel 2.2 berikut:

Tabel 2.2 Pengelompokan Atribut pada NSL KDD CUP

Tipe	Fitur (atribut)
Nominal	<i>Protocol_type</i> (2), <i>Services</i> (3), <i>Flag</i> (4)
Binary	<i>Land</i> (7), <i>logged_in</i> (12), <i>root_shell</i> (14), <i>su_attempted</i> (15), <i>is_host_login</i> (21), <i>is_guest_login</i> (22)
Numeric	<i>Duration</i> (1), <i>src_bytes</i> (5), <i>dst_bytes</i> (6), <i>wrong_fragment</i> (8), <i>urgent</i> (9), <i>hot</i> (10), <i>num_failed_logins</i> (11), <i>num_compromised</i> (13), <i>num_root</i> (16),

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

```
num_file_creations(17),
num_shells(18),
num_access_files(19),
num_outbound_cmds(20), count(23)
srv_count(24), serror_rate(25),
srv_serror_rate(26), error_rate(27),
srv_error_rate(28), same_srv_rate(29)
diff_srv_rate(30),
srv_diff_host_rate(31),
dst_host_count(32),
dst_host_srv_count(33),
dst_host_same_srv_rate(34),
dst_host_diff_srv_rate(35),
dst_host_same_src_port_rate(36),
dst_host_srv_diff_host_rate(37),
dst_host_serror_rate(38),
dst_host_srv_serror_rate(39),
dst_host_error_rate(40),
dst_host_srv_error_rate(41)
```

Sumber : (Dhanabal dan Shantharajah, 2015)

Jenis serangan pada NSL KDD CUP 1999 dibagi menjadi 4 kategori yakni DoS (*Denial of Service*), *Probe*, U2R (*User to Root*), R2L (*Remote to Local*) (Dhanabal dan Shantharajah, 2015).

DoS (*Denial of Service*) ialah tipe serangan yang membebani sumber daya komputer. Sehingga komputer target mengalami *crash* dan tidak dapat terhubung ke koneksi normal bahkan mengakibatkan pengguna tidak dapat mengakses komputer yang diserang.

Probe ialah tipe serangan yang bertujuan untuk mendapatkan informasi tentang status jaringan komputer dengan cara melakukan pemindaian terhadap komputer yang berada pada lingkup suatu jaringan komputer. Informasi tersebut digunakan oleh penyerang untuk pemetaan jaringan yang bisa dijadikan acuan untuk penyerangan berikutnya.

U2R (*User to Root*) ialah tipe serangan yang berusaha untuk menjebol akses root/admin ke komputer target dengan cara melakukan eksploitasi terhadap celah keamanan sistem. Serangan U2R dapat dilancarkan setelah penyerang mendapatkan akses user *normal* terhadap sistem dengan cara *sniffing* ataupun *social engineering*.

R2L (*Remote to Local*) ialah tipe serangan yang bertujuan untuk mendapatkan akses sebagai pengguna sistem. R2L dilakukan oleh penyerang yang

memiliki akses ke sistem serta melakukan eksploitasi untuk mendapatkan akses lokal.

Tabel 2.3 Daftar Kelas Serangan pada NSL KDD CUP

<i>Attack Class</i>	<i>Attack Type</i>
DoS	<i>Back, Land, Neptune, Pod, Smurf, Teardrop, Apache2, Udpstorm, Processtable, Worm</i>
Probe	<i>Satan, Ipsweep, Nmap, Portsweep, Mscan, Saint</i>
R2L	<i>Guess_Password, Ftp_write, Imap, Phf, Multihop, Warezmater, Warezclient, Spy, Xlock, Xsnoop, Snmpguess, Snmpgetattack, Httptunnel, Sendmail, Named</i>
U2R	<i>Buffer_overflow, Loadmodule, Rootkit, Perl, Sslattack, Xterm, Ps</i>

Sumber : (Dhanabal dan Shantharajah, 2015)

2.7 Evaluasi

Evaluasi merupakan kunci utama dari pembuatan aplikasi berbasis *data mining*. Kemampuan dari suatu model klasifikasi dapat diukur dengan tingkat akurasi berdasarkan *confusion matrix*.. Berikut ini ialah bentuk tabel dari *confusion matrix* yang dapat dilihat pada Tabel 2.4 berikut:

Tabel 2.4 Confusion Matrix

Kelas Prediksi	Kelas Aktual	
	Positives (+)	Negatives (-)
Positives (+)	TP	FN
Negatives (-)	FP	TN

TP dan TN memberikan informasi ketika *classifier* ialah benar, sedangkan FP dan FN memberikan informasi ketika *classifier* ialah salah. Serta akurasi ialah persentase dari data yang diprediksi secara benar. Berikut ini adalah rumus yang digunakan untuk mengecek akurasi:

$$Akurasi = \frac{(TP+TN)}{(TP+TN+FP+FN)} \dots \dots \dots (2.10)$$

Dimana :

TP : *True Positives*, jumlah data dengan kelas positif yang diklasifikasikan positif

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

TN : *True Negatives*, jumlah data dengan kelas negatif yang diklasifikasikan negatif

FP : *False Postives*, jumlah data dengan kelas positif yang diklasifikasikan negatif

FN : *False Negatives*, yakni jumlah data dengan kelas negatif yang diklasifikasikan positif

Kajian Penelitian Terkait

Berikut beberapa penelitian terkait terhadap metode yang digunakan pada penelitian ini, dapat dilihat pada Tabel 2.5 berikut.

Tabel 2.5 Penelitian Terkait

No	Penulis	Tahun	Judul	Hasil
1	Diina Itsna Annisa, Rudy Ariyanto, Ariadi Tri Retno Hayati Ririd	2016	Klasifikasi Kehamilan Beresiko Dengan Menggunakan Metode <i>K-Nearest Neighbor</i> (Studi Kasus Dinas Kesehatan Kabupaten Malang)	Hasil pengujian pada penelitian ini diperoleh sebesar 93% dengan membandingkan data yang diperoleh dari dinas kesehatan terhadap sistem dengan menggunakan metode validasi. Nilai K yang digunakan ialah 5.
2	Kevin Martha Rasepta	2016	Klasifikasi Status Gizi Balita Menggunakan Metode <i>Modifiked K-Nearest Neighbor</i>	Hasil akurasi tertinggi yang diperoleh pada penelitian ini ialah sebesar 90% pada skema 90:10 dan 80:20 dengan nilai $k=1$ hingga $k=3$
	Fakihatun Wafiyah, Nurul Hidayat dan Rizal Setya Perdana	2017	Implementasi Algoritma <i>Modified K-Nearest Neighbour</i> untuk klasifikasi Penyakit Demam	Nilai rata-rata akurasi pengujian pengaruh nilai K terhadap akurasi ialah 88,55% dengan akurasi tertinggi pada pengujian nilai $K = 3$ dengan akurasi sebesar 94,95%
	Dian Eka Ratnawati, Putra Pandu Adikara dan Silvia Ikmalia Fernanda	2017	Identifikasi Penyakit Diabetes Mellitus Menggunakan Metode <i>Modified K-Nearest Neighbour</i>	Berdasarkan pengujian yang telah dilakukan diperoleh hasil akurasi terbaik sebesar 93,33% dengan menggunakan nilai $K=3$



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tri Halomon Simanjutak, Wayan Firdaus Mahmudy dan Sutrisno	2017	Implementasi <i>Modified K-Nearest Neighbor</i> dengan otomatisasi Nilai K Pada Pengklasifikasian Penyakit Tanaman Kedelai.	Hasil akurasi tertinggi pada pengujian ini sebesar 100% dengan nilai K=1 dan rata-rata akurasi dari 5 percobaan sebesar 98,83%
Aldion Cahya Imanda, Nurul Hidayat dan Muhammad Tanzil Furqon	2017	Klasifikasi Kelompok Varietas Unggul Padi Menggunakan <i>Modified K-Nearest Neighbor</i>	Hasil akurasi rata-rata yang diperoleh pada penelitian ini dengan nilai tertinggi ialah 79,96% dan nilai terendah ialah 51,2%
Tito Yuwono, Annafi' Franz dan Izzati Muhimmah	2018	<i>Design of Smart Electrocardiography (ECG) Using Modified K-Nearest Neighbor (MKNN)</i>	Akurasi yang diperoleh pada metode ini ialah sebesar 71.20% dengan nilai K=3 terhadap 3 kelas serangan jantung.

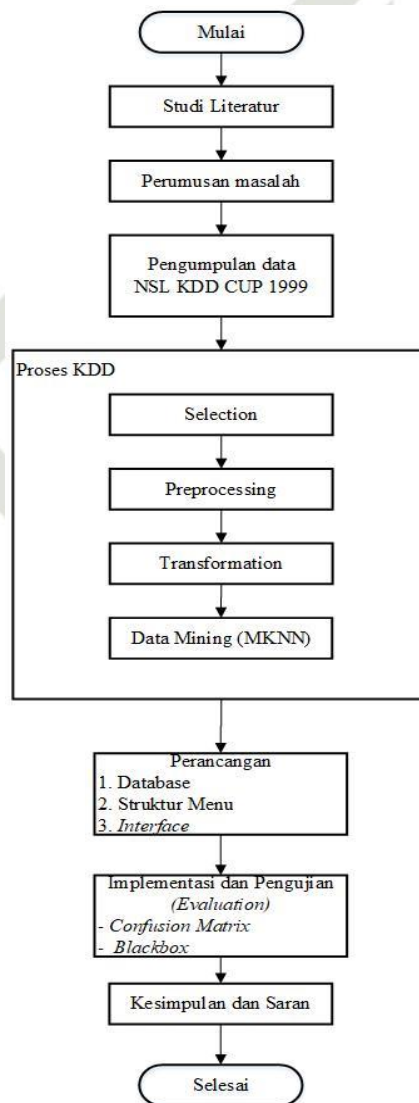
Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB III

METODOLOGI PENELITIAN

Metodologi penelitian ialah acuan atau tahapan-tahapan yang akan dilakukan selama penelitian agar penelitian dapat berjalan sesuai prosedur dan mencapai hasil yang maksimal. Metodologi penelitian pada tugas akhir ini dapat dilihat pada Gambar 3.1 berikut:



Gambar 3.1 Metodologi Penelitian



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3.1 Studi Literatur

Studi literatur ialah langkah yang digunakan untuk mengumpulkan data serta informasi dengan cara membaca referensi yang berasal dari buku dan jurnal yang berhubungan dengan penelitian pengklasifikasian *dataset* dengan menggunakan metode MKNN.

3.2 Perumusan Masalah

Pada tahap perumusan masalah dilakukan perumusan dan memberikan kesimpulan terhadap permasalahan yang telah dicantumkan dalam latar belakang pada pembahasan sebelumnya. Perumusan masalah ini dibuat dengan tujuan agar dapat lebih memahami penelitian yang dilakukan, yaitu mengenai permasalahan yang akan dibahas dalam penelitian, mencari tujuan dan membuat batasan masalah.

3.3 Pengumpulan Data

Tahapan pengumpulan data adalah tahapan awal yang dilakukan sebelum dilakukannya penelitian. Proses pengumpulan data pada penelitian ini ialah dengan mendapatkan *dataset* NSL KDD CUP 1999 yang memiliki 41 parameter (atribut) dan 4 kategori serangan.

3.4 Proses KDD (*Knowledge Discovery in Database*)

Pada proses ini bertujuan untuk menemukan pengetahuan ataupun informasi dari *dataset* NSL KDD CUP 1999 yakni berupa klasifikasi serangan. Pengetahuan tersebut dihasilkan dengan melewati beberapa tahapan ataupun prosedur berikut.

3.4.1 Selection

Pada proses ini dilakukan pemilihan himpunan data, menciptakan himpunan data target, atau memfokuskan pada *subset* variabel, (sampel data) pada *dataset* NSL KDD CUP 1999. Pada *dataset* ini terdapat 41 atribut. Dari 41 atribut tersebut akan diseleksi dan disimpan dalam suatu berkas yang terpisah dari basis data operasional.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3.4.2 Pre Processing

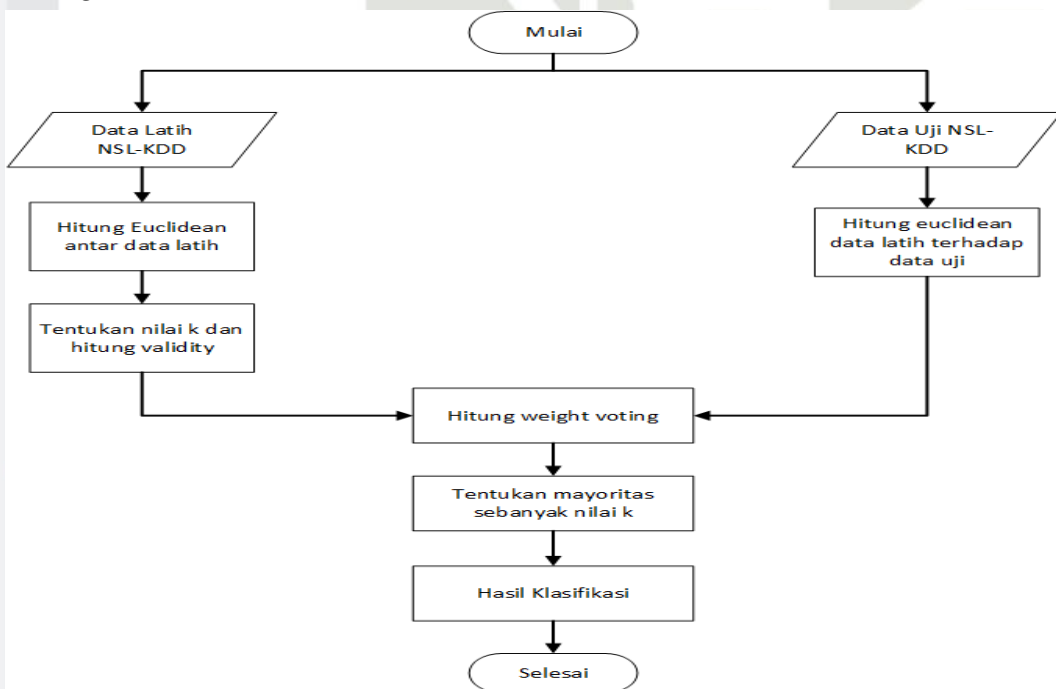
Pada proses ini dilakukan pembersihan data NSL-KDD. Proses pembersihan meliputi penghapusan data duplikat, penghapusan data yang terdapat *missing value* dan pengecekan serta penghapusan data *outlier*.

3.4.3 Transformation

Pada proses ini akan dilakukan pentransformasian atau penggabungan *dataset* NSL KDD CUP 1999. Cara yang dilakukan pada proses ini ialah dengan menggunakan normalisasi Persamaan (2.1). Pada proses normalisasi nilai pada data atribut akan dibuat dalam skala 0 hingga 1 sehingga sebaran data tidak terlalu jauh. Proses normalisasi dilakukan dengan menggunakan persamaan *min-max normalization*, dimana x' adalah hasil nilai normalisasi, $\min(x)$ adalah nilai minimum pada atribut x dan $\max(x)$ ialah nilai maksimum pada atribut x (Larose, 2005).

3.4.4 Data Mining

Pada proses ini ialah dilakukannya pengklasifikasian data serangan jaringan komputer dengan menggunakan metode *Modified K-Nearest Neighbor*. Berikut ialah *flowchart* proses dari klasifikasi MKNN (*Modified K-Nearest Neighbor*):



Gambar 3.2 Flowchart MKNN

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Berdasarkan Gambar 3.2 di atas proses MKNN pada penelitian ini melalui beberapa tahapan yaitu penginputan *dataset* NSL KDD CUP 1999 yang telah dinormalisasi menggunakan Persamaan (2.1), melakukan perhitungan jarak *euclidean* antara data latih menggunakan Persamaan (2.4), melakukan perhitungan validitas pada data latih menggunakan Persamaan (2.5), melakukan perhitungan jarak *euclidean* antara data uji dan data latih menggunakan Persamaan (2.4), perhitungan bobot/*weight voting* pada *dataset* menggunakan Persamaan (2.9) dan kemudian didapat hasil klasifikasi setelah ditentukan mayoritas kelas data yang meliputi normal, serangan probe, serangan Denial of Service (DOS), serangan remote to local (R2L) dan serangan user to root (U2R).

3.5 Perancangan

Pada tahapan perancangan ini terdapat 3 tahapan yang dilakukan yakni:

1. Perancangan basis data

Pada tahapan ini merupakan tahap perancangan tabel dan atribut yang dibutuhkan oleh sistem.

2. Perancangan struktur menu

Pada tahapan perancangan struktur menu akan memberikan gambaran terhadap menu-menu serta fitur sistem yang akan dibangun.

Perancangan *interface* (antar muka)

Pada tahapan perancangan *interface* akan memberikan gambaran atau konsep awal desain antarmuka sistem yang akan dibangun.

3.6 Implementasi dan Pengujian (*Evaluation*)

Implementasi pada tahap ini merupakan tahap pembuatan dari aplikasi berdasarkan rancangan yang telah dibuat sebelumnya. Sedangkan pengujian merupakan tahap *evaluation*, untuk menerjemahkan pola-pola yang dihasilkan dari tahapan data mining. Pengujian dilakukan dengan menggunakan dua cara, yakni:

3.6.1 Pengujian Akurasi

Pengujian akurasi dilakukan dengan menggunakan Persamaan (2.10). Pada *Confusion matrix* setiap kolom pada matriks ialah contoh dari kelas prediksi, dan

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

tiap baris mewakili kejadian pada kelas yang sebenarnya. Selain itu, pada *confusion matrix* berisi informasi aktual dan prediksi pada sistem klasifikasi.

3.6.2 Black Box

Pada pengujian blackbox akan memastikan semua fungsi berjalan dengan baik guna mendapatkan keakurasian aplikasi yang telah dibangun. Pengujian ini yang berfokus pada spesifikasi fungsional dari perangkat lunak, tester dapat mendefinisikan kumpulan kondisi input dan melakukan pengetesan pada spesifikasi fungsional program.

3.7 Kesimpulan dan Saran

Pada tahapan ini dirumuskan kesimpulan terkait hasil dari pengujian yang telah dilakukan. Selain itu juga diberikan saran untuk penyempurnaan serta pengembangan hasil penelitian ini untuk selanjutnya

BAB IV

ANALISA DAN PERANCANGAN

4.1 Analisa Data

Analisa data merupakan tahapan yang menjadi acuan dalam membangun sebuah sistem karena hasil analisa data akan menentukan hasil dari sistem yang akan dibangun. Data yang digunakan dalam penelitian ini, menggunakan data serangan dari NSL-KDD yang merupakan versi pembaharuan dari data serangan KDD *Cup*, dengan jumlah data latih yakni 125.973 dan data uji berjumlah 22.544. Yang terdiri atas 41 variabel dan 5 kelas yakni normal, DoS, U2R, R2L dan *Probe*. Pada tahapan analisa ini pengolahan data dilakukan dengan menggunakan proses KDD (*Knowledge Discovery Data*) yang memiliki 5 tahapan yakni *data selection*, *pre-processing data*, transformasi data, *data mining* serta *evaluation*.

4.2 Tahapan KDD (*Knowledge Discovery Data*)

Pada tahapan ini sebelum data masukan di integrasikan pada sistem, data di analisa terlebih dahulu melewati tahapan-tahapan yang terdapat pada proses KDD untuk keperluan proses pengklasifikasian serangan. Berikut tahapan dari proses KDD:

4.2.1 Data Selection

Data masukan yang akan di olah pada penelitian ini menggunakan metode KNN dengan menggunakan data NSL-KDD yang memiliki 41 variabel dan 5 kelas. Pada penelitian ini akan menggunakan 29 variabel, pemilihan 29 variabel ini berdasarkan penelitian sebelumnya yang dilakukan oleh (Ingre dan Yadav, 2015) yang diperoleh dengan menggunakan metode seleksi fitur yang terdiri dari *Information Gain Attribute Evaluation*, *Gain Ratio Attribute* dan *Correlation Attribute Evaluation* dengan menyisakan 33 variabel. Serta terdapat 4 variabel yang dihapus karena keseluruhan datanya bernilai 0 yakni pada variabel *land*, *wrong_fragment*, *num_failed_login* dan *root_shell* (Ingre dan Yadav, 2015). Berikut ialah variabel yang digunakan pada penelitian ini pada Tabel 4.1:

Tabel 4.1 Data Variabel

No	Variabel yang digunakan	Keterangan
1	Duration	A1
2	Protocol_type	A2
3	Service	A3
4	Flag	A4
5	Src_bytes	A5
6	Dst_bytes	A6
7	Hot	A7
8	Logged_in	A8
9	Num_compromised	A9
10	Num_root	A10
11	Is_guest_login	A11
12	Count	A12
13	Srv_count	A13
14	Serror_rate	A14
15	Srv_serror_rate	A15
16	Rerror_rate	A16
17	Srv_rerror_rate	A17
18	Same_srv_rate	A18
19	Diff_srv_rate	A19
20	Srv_diff_host_rate	A20
21	Dst_host_srv_count	A21
22	Dst_host_same_srv_rate	A22
23	Dst_host_diff_srv_rate	A23
24	Dst_host_same_src_port_rate	A24
25	Dst_host_srv_diff_host_rate	A25
26	Dst_host_serror_rate	A26
27	Dst_host_srv_serror_rate	A27
28	Dst_host_rerror_rate	A28
29	Dst_host_srv_rerror_rate	A29

4.2.2 Preprocessing data

Pada tahapan ini menentukan jumlah data yang akan diproses terhadap pembelajaran metode MKNN nantinya dengan membersihkan data ataupun menghapus data yang bersifat redundan, *ouliter* serta mengisi nilai-nilai variabel

yang hilang (*missing value*). Berikut ini ialah data serangan NSL-KDD yang dapat dilihat pada Tabel 4.2 :

Tabel 4.2 Data NSL KDD Variabel 1-29

[illegible]

- | Hak Cipta Dilindungi | Ya | Tidak | Jumlah |
|---|----|-------|--------|
| 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber. | 0 | 0 | 0 |
| a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah. | 0 | 0 | 0 |
| b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau. | 0 | 0 | 0 |
| 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau. | 0 | 0 | 0 |



Hak Cipta Dilindungi Undang-Undang

Dua	A22	A23	A24	A25	A26	A27	A28	A29	Kelas
1	0	2	2	0	0	0	0	1	normal
2	0	13	1	0	0	0	0	0.08	normal
3	0	123	6	1	1	0	0	0.05	neptune
4	0	5	5	0.2	0.2	0	0	1	normal
5	0	30	32	0	0	0	0	1	normal
6	0	121	19	0	0	1	1	0.16	neptune
7	0	166	9	1	1	0	0	0.05	neptune
8	0	117	16	1	1	0	0	0.14	neptune
9	0	270	23	1	1	0	0	0.09	neptune
10	0	133	8	1	1	0	0	0.06	neptune
...
125.973	0	1	1	0	0	0	0	1	normal

1. Diarang mengutip sebagian atau seluruhnya tulisan tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan satu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Diarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Berikut ini ialah proses tahapan-tahapan *pre processing* pada data NSL-KDD :

Missing Value

Pada tahapan ini melakukan pembuangan data terhadap variabel yang kosong atau tidak terdapat nilai. Proses pengecekan *missing value* dilakukan dengan menggunakan *tool* RapidMiner. Setelah dilakukan nya proses pengecekan pada data, **tidak ditemukan** adanya data yang bersifat *missing value*.

2. Penghapusan *Data Duplicate*

Pada tahapan ini dilakukan pengecekan serta penghapusan data duplikat yang terdapat pada serangan NSL-KDD. Proses pengecekan dilakukan dengan menggunakan *tool* RapidMiner. Setelah dilakukan proses pengecekan pada data, **tidak ditemukan adanya** data yang *duplicate*.

3. Penghapusan *Data Outlier*

Pada tahapan ini dilakukan pengecekan data outlier tiap variabel yang digunakan. Proses pengecekan dan penghapusan data outlier dilakukan dengan menggunakan aplikasi jupyter yang berbasis bahasa python. Data yang dihapus ialah data yang memiliki *z-score* > 2,5 dan *z-score* < -2,5 (Ghozali, 2006). Standar deviasi SD(x) diperoleh dengan menggunakan Persamaan (2.3) dan Nilai *z-score* diperoleh dengan menggunakan Persamaan (2.2). Berikut perhitungan *z-score* berdasarkan Tabel 4.2:

Pencarian *outlier* untuk variabel *duration* :

Nilai x data pertama = 0

Nilai *mean(x)* = 287,1447

$$\begin{aligned} \text{Nilai SD}(x) &= \sqrt{\frac{(125.973)(864917794017) - (1308447802935730)}{(125.973) (125.972)}} \\ &= 2604,515 \end{aligned}$$

$$X' = \frac{0 - 287,1447}{2604,515} = -0,11025$$

Berikut ialah proses pencarian serta penghapusan data outlier dengan menggunakan aplikasi jupyter :

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- Pertama menginputkan data latih ke dalam variabel *datazsc* dan dihitung nilai *z-score* tiap variabel dengan menggunakan perintah *stats.zscore* untuk mencari nilai *z-score* tiap data. Setelah nilai *z-score* didapatkan, proses selanjutnya dilakukan penyaringan nilai *z-score* tiap data dengan menggunakan perintah *np.nonzero* (*datazsc* > 2,5 dan *datazsc* < -2,5) dan disimpan pada variabel *filter_outlier*.

```
In [8]: # Menghitung z-score data training dan menyimpannya ke variabel datazsc
datazsc = stats.zscore(taproc)
# Menyaring z-score setiap field yang nilainya lebih dari 2.5 atau kecil dari -2.5
filter_outlier = np.nonzero((datazsc > 2.5) | (datazsc < -2.5))
```

Gambar 4.1 Pencarian dan Penyaringan Nilai Z-score data latih

- Selanjutnya menginputkan data uji ke dalam variabel *datazsc* dan dihitung nilai *z-score* tiap variabel dengan menggunakan perintah *stats.zscore* untuk mencari nilai *z-score* tiap data. Serta dilakukan penyaringan data latih dengan *z-score* > 2,5 dan *z-score* < -2,5

```
In [10]: #Menghitung z-score data testing dan disimpan ke variabel datazsc
datestzsc = stats.zscore(testproc)
# Menyaring z-score setiap field yang nilainya lebih dari 2.5 atau kecil dari -2.5
filter_outlier_test = np.nonzero((datestzsc > 2.5) | (datestzsc < -2.5))
```

Gambar 4.2 Penyaringan dan Penyaringan Nilai Z-score data uji

- Berikut ialah hasil pencarian nilai *zscore* dari tiap variabel yang termasuk kategori data outlier pada data latih.

In [13]: # Z-Score setiap atribut data training yang terkena outlier
pd.DataFrame(datazsc).loc[np.unique(filter_outlier[0])]

Out[13]:

	0	1	2	3	4	5	6	7	8	9 ...	27	28	29	30	
1	-0.007737	-0.004919	-0.014089	-0.089486	-0.007736	-0.095076	-0.027023	-0.809262	-0.011664	-0.036652	...	-1.035688	-1.161030	2.736852	2.367773
5	-0.007762	-0.004919	-0.014089	-0.089486	-0.007736	-0.095076	-0.027023	-0.809262	-0.011664	-0.036652	...	-0.873089	-1.005110	-0.068553	-0.480191
10	-0.007762	-0.004919	-0.014089	-0.089486	-0.007736	-0.095076	-0.027023	-0.809262	-0.011664	-0.036652	...	-0.936322	-1.049659	-0.068553	-0.480191
13	-0.007705	-0.004919	-0.014089	-0.089486	-0.007736	-0.095076	-0.027023	1.235694	-0.011664	-0.036652	...	-0.864056	1.066401	-0.439078	2.756091
17	-0.007759	-0.004919	-0.014089	-0.089486	-0.007736	-0.095076	-0.027023	-0.809262	-0.011664	-0.036652	...	-0.900189	1.066401	-0.439078	2.756091
23	-0.007737	-0.004893	-0.014089	-0.089486	-0.007736	-0.095076	-0.027023	-0.809262	-0.011664	-0.036652	...	-1.035688	-1.161030	4.060156	2.756091
25	-0.007688	-0.001332	-0.014089	-0.089486	-0.007736	-0.095076	-0.027023	1.235694	0.113639	-0.036652	...	-0.818890	-0.938287	-0.174417	-0.480191
30	-0.007761	-0.004919	-0.014089	-0.089486	-0.007736	-0.095076	-0.027023	-0.809262	-0.011664	-0.036652	...	-0.565959	1.066401	-0.439078	2.756091
33	-0.007762	-0.004919	-0.014089	-0.089486	-0.007736	-0.095076	-0.027023	-0.809262	-0.011664	-0.036652	...	-1.035688	-1.161030	1.201819	0.425961
42	-0.007491	-0.004826	-0.014089	-0.089486	-0.007736	-0.095076	-0.027023	1.235694	-0.011664	-0.036652	...	0.283165	0.643189	-0.333214	-0.447831
43	-0.007737	-0.004893	-0.014089	-0.089486	-0.007736	-0.095076	-0.027023	-0.809262	-0.011664	-0.036652	...	-1.026654	-1.138756	3.054445	2.723721

Gambar 4.3 Data Outlier

- Berikut ialah *z-score* dari data training yang merupakan data outlier yang telah diekspor dalam bentuk csv. Terlihat pada Gambar 4.4 angka yang

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumpukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

diberi warna merah merupakan data z -score yang memiliki nilai $zscore > 2,5$.

Zduration	Zsrc_byte	Zdst_byte	Zhot	Zlogged_i	Znum_cor	Znum_roc	Zis_guest	Zcount	Zsrv_coun	Zerror_ra	Zsrv_serrc	Zerror_ra	Zsrv_rerrc
-0.11025	-0.00492	-0.09508	-0.80926	-0.01166	-0.01239	-0.01861	-0.09753	-0.62098	-0.36811	-0.63721	-0.63193	-0.37436	-0.37443
-0.11025	-0.00492	-0.09508	-0.80926	-0.01166	-0.01239	-0.01861	-0.09753	0.322182	-0.1203	-0.63721	-0.63193	2.746403	2.715365
-0.11025	-0.00492	-0.09508	-0.80926	-0.01166	-0.01239	-0.01861	-0.09753	1.055754	-0.21667	-0.63721	-0.63193	2.746403	2.715365
-0.11025	-0.00492	-0.09508	1.235694	-0.01166	-0.01239	-0.01861	-0.09753	-0.71705	-0.35434	-0.63721	-0.63193	-0.37436	-0.37443
-0.11025	-0.00492	-0.09508	-0.80926	-0.01166	-0.01239	-0.01861	-0.09753	-0.72578	-0.36811	-0.63721	-0.63193	-0.37436	-0.37443
2.042559	-0.00489	-0.09508	-0.80926	-0.01166	-0.01239	-0.01861	-0.09753	-0.72578	-0.36811	-0.63721	-0.63193	-0.37436	-0.37443
-0.11025	-0.00492	-0.09508	-0.80926	-0.01166	-0.01239	-0.01861	-0.09753	-0.72578	-0.21667	-0.63721	-0.63193	-0.37436	-0.37443
-0.11025	-0.00492	-0.09508	-0.80926	-0.01166	-0.01239	-0.01861	-0.09753	-0.71705	-0.36811	-0.63721	-0.63193	2.746403	2.715365
-0.10948	-0.00483	-0.09508	1.235694	-0.01166	-0.01239	-0.01861	-0.09753	-0.72578	-0.35434	-0.63721	-0.63193	-0.37436	-0.37443
3.365267	-0.00489	-0.09508	-0.80926	-0.01166	-0.01239	-0.01861	-0.09753	-0.69958	-0.35434	-0.63721	-0.63193	-0.37436	-0.37443
-0.11025	-0.00492	-0.09508	1.235694	-0.01166	-0.01239	-0.01861	-0.09753	-0.71705	-0.35434	-0.63721	-0.63193	-0.37436	-0.37443
-0.11025	-0.00492	-0.09508	-0.80926	-0.01166	-0.01239	-0.01861	-0.09753	-0.2018	-0.32681	-0.63721	-0.63193	2.746403	2.715365
-0.11025	-0.00492	-0.09508	-0.80926	-0.01166	-0.01239	-0.01861	-0.09753	-0.72578	-0.17537	-0.63721	-0.63193	-0.37436	-0.37443
-0.11025	-0.00491	-0.09508	-0.80926	-0.01166	-0.01239	-0.01861	-0.09753	0.557973	2.757081	-0.63721	-0.63193	-0.37436	-0.37443
-0.11025	-0.00492	-0.09508	-0.80926	-0.01166	-0.01239	-0.01861	-0.09753	1.064487	-0.32681	-0.63721	-0.63193	2.746403	2.715365
-0.11025	-0.00492	-0.09508	-0.80926	-0.01166	-0.01239	-0.01861	-0.09753	0.793764	-0.36811	-0.41322	-0.63193	2.403119	2.715365
-0.11025	-0.00492	-0.09508	-0.80926	-0.01166	-0.01239	-0.01861	-0.09753	-0.18433	-0.2442	-0.63721	-0.63193	2.746403	2.715365
-0.11025	-0.00492	-0.09508	-0.80926	-0.01166	-0.01239	-0.01861	-0.09753	-0.72578	-0.36811	-0.63721	-0.63193	2.746403	2.715365

Gambar 4.4 Data Outlier Csv

Jumlah data latih dan data uji yang tersisa setelah dilakukan proses pengecekan serta penghapusan data *outlier* ialah 88.724 data latih dan 14.537 data uji.

4.2.3 Transformasi Data

Pada tahapan transformasi data dilakukan pengubahan data teks menjadi bentuk numerik pada atribut data NSL-KDD. Atribut yang ditransformasi ialah kelas, *protocol*, *service* dan *flag*. Proses transformasi ini dilakukan berdasarkan penelitian sebelumnya oleh (Ibrahim, Basheer, dan Mahmod, 2013). Berikut ialah hasil proses transformasi atribut kelas pada Tabel 4.3:

Tabel 4.3 Transformasi Kelas

Atribut	Transformasi
Normal	1
Dos	2
R2L	3
U2R	4
Probe	5

Berikut ialah hasil proses transformasi pada atribut *protocol*, *service* dan *flag* pada Tabel 4.4:

Tabel 4.4 Transformasi Protocol, Service dan Flag

Atribut	Transformasi
<i>Protocol</i>	tcp:1 udp:2 icmp:3
<i>Service</i>	http:1, private:2, domain_u:3, smtp:4, ftp_data:5, eco_i:6, other:7, ecr_i:8, telnet:9, finger:10, ftp:11, auth:12, Z39_50:13, uucp:14, courier:15, bgp:16, whois:17, uucp_path:18,



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

	iso_tsap:19, time:20, imap4:21, nnsp:22, vmnet:23, urp_i:24, domain:25, ctf:26, csnet_ns:27, supdup:28, discard:29, http_443:30, daytime:31, gopher:32, efs:33, systat:34, link:35, exec:36, hostnames:37, name:38, mtp:39, echo:40, klogin:41, login:42, ldap:43, netbios_dgm:44, sunrpc:45, netbios_ssn:46, netstat:47, netbios_ns:48, ssh:49, kshell:50, nntp:51, pop_2:52, sql_net:53, IRC:54, ntp_u:55, rje:56, pop_3:57, remote_job:58, X11:59, printer:60, shell:61, urh_i:62, red_i:63, tim_i:64, pm_dump:65, tftp_u:66, aol:67, harvest:68, http_8001:69, http_2784:70
Flag	SF:1, S0:2, REJ:3, RSTR:4, RSTD:5, S1:6, SH:7, S2:8, RSTOS0:9, S3:10, OTH:11, RSTO:12

Berikut ialah data yang telah ditransformasi yang dapat dilihat pada Tabel 4.5 berikut:



Tabel 4.5 Data Transformasi

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20	A21
1	0	1	5	1	491	0	0	0	0	0	0	0	2	2	0	0	0	0	1	0	0
2	0	1	2	2	0	0	0	0	0	0	0	0	123	6	1	1	0	0	0.05	0.07	0
3	0	1	1	1	232	8153	0	1	0	0	0	0	5	5	0.2	0.2	0	0	1	0	0
4	0	1	1	1	199	420	0	1	0	0	0	0	30	32	0	0	0	0	1	0	0.09
5	0	1	2	2	0	0	0	0	0	0	0	0	166	9	1	1	0	0	0.05	0.06	0
6	0	1	2	2	0	0	0	0	0	0	0	0	117	16	1	1	0	0	0.14	0.06	0
7	0	1	58	2	0	0	0	0	0	0	0	0	270	23	1	1	0	0	0.09	0.05	0
8	0	1	2	2	0	0	0	0	0	0	0	0	133	8	1	1	0	0	0.06	0.06	0
9	0	1	2	2	0	0	0	0	0	0	0	0	199	3	1	1	0	0	0.02	0.06	0
10	0	1	1	1	287	2251	0	1	0	0	0	0	3	7	0	0	0	0	1	0	0.43
...
88	0	1	5	1	151	0	0	1	0	0	0	0	1	1	0	0	0	0	1	0	0

- Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

Data ke	A22	A23	A24	A25	A26	A27	A28	A29	Kelas
1	25	0.17	0.03	0.17	0	0	0	0	1
2	26	0.1	0.05	0	0	1	1	0	2
3	255	1	0	0.03	0.04	0.03	0.01	0.01	1
4	255	1	0	0	0	0	0	0	1
5	9	0.04	0.05	0	0	1	1	0	2
6	15	0.06	0.07	0	0	1	1	0	2
7	23	0.09	0.05	0	0	1	1	0	2
8	13	0.05	0.06	0	0	1	1	0	2
9	13	0.05	0.07	0	0	1	1	0	2
10	219	1	0	0.12	0.03	0	0	0	1
...
88.724	77	0.3	0.03	0.3	0	0	0	0	1

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumpukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Selanjutnya dilakukan proses normalisasi dengan tujuan mengatasi sebaran data agar nilai dari masing-masing atribut/varibel tidak terlalu jauh. Nilai data pada setiap atribut akan diubah pada rentang 0-1. Proses normalisasi dilakukan dengan menggunakan Persamaan (2.1). Berdasarkan Tabel 4.5 proses perhitungan normalisasi dapat dijabarkan sebagai berikut:

- a. Normalisasi untuk variabel *duration*:

Nilai x untuk data pertama = 0

Nilai $\min(x)$ untuk nilai terendah = 0

Nilai $\max(x)$ untuk nilai tertinggi = 6552

$$X = \frac{0 - 0}{6552 - 0} = 0$$

- b. Normalisasi untuk variabel *protocol*:

Nilai x untuk data pertama = 1

Nilai $\min(x)$ untuk nilai terendah = 1

Nilai $\max(x)$ untuk nilai tertinggi = 3

$$X = \frac{1 - 1}{3 - 1} = 0$$

- c. Normalisasi untuk variabel *service*:

Nilai x untuk data pertama = 5

Nilai $\min(x)$ untuk nilai terendah = 1

Nilai $\max(x)$ untuk nilai tertinggi = 66

$$X = \frac{5 - 1}{66 - 1} = 0,06154$$

Perhitungan diatas dilakukan pada setiap data penelitian. Berikut ini ialah data hasil normalisasi yang dapat dilihat pada Tabel 4.6 berikut:

UIN SUSKA RIAU

Tabel 4.6 Hasil Normalisasi

Data	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19
1	0	0	0.0615	0	3.53×10^{-5}	0	0	0	0	0	0	0	0.0066	0.0079	0	0	0	0	1
2	0	0	0.0154	0.0909	0	0	0	0	0	0	0	0	0.4046	0.0238	1	1	0	0	0.05
3	0	0	0	0	1.67×10^{-5}	0.0016	0	1	0	0	0	0	0.0164	0.0198	0.2	0.2	0	0	1
4	0	0	0	0	1.43×10^{-5}	8.18×10^{-5}	0	1	0	0	0	0	0.0987	0.1270	0	0	0	0	1
5	0	0	0.0154	0.0909	0	0	0	0	0	0	0	0	0.5461	0.0357	1	1	0	0	0.05
6	0	0	0.0154	0.0909	0	0	0	0	0	0	0	0	0.3849	0.0635	1	1	0	0	0.14
7	0	0	0.8769	0.0909	0	0	0	0	0	0	0	0	0.8882	0.0913	1	1	0	0	0.09
8	0	0	0.0154	0.0909	0	0	0	0	0	0	0	0	0.4375	0.0317	1	1	0	0	0.06
9	0	0	0.0154	0.0909	0	0	0	0	0	0	0	0	0.6546	0.0119	1	1	0	0	0.02
10	0	0	0	0	2.06×10^{-5}	0.0004	0	1	0	0	0	0	0.0099	0.0278	0	0	0	0	1
11	0	0	0.5692	0.0909	0	0	0	0	0	0	0	0	0.7664	0.0040	1	1	0	0	0
12	0	0	0.7231	0.0909	0	0	0	0	0	0	0	0	0.3158	0.0635	1	1	0	0	0.17
13	0	0	0	0	2.16×10^{-5}	0.0027	0	1	0	0	0	0	0.0263	0.0357	0	0.11	0	0	1
14	0	0	0	0	1.68×10^{-5}	0.0001	0	1	0	0	0	0	0.0099	0.0119	0	0	0	0	1
15	0	0	0	0	2.47×10^{-5}	0.0002	0	1	0	0	0	0	0.0296	0.0397	0	0	0	0	1
16
88.74	0	0	0.0615	0	1.09×10^{-5}	0	0	1	0	0	0	0	0.0033	0.0040	0	0	0	0	1

1. Diarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Diarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Diarangi mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Diarangi mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

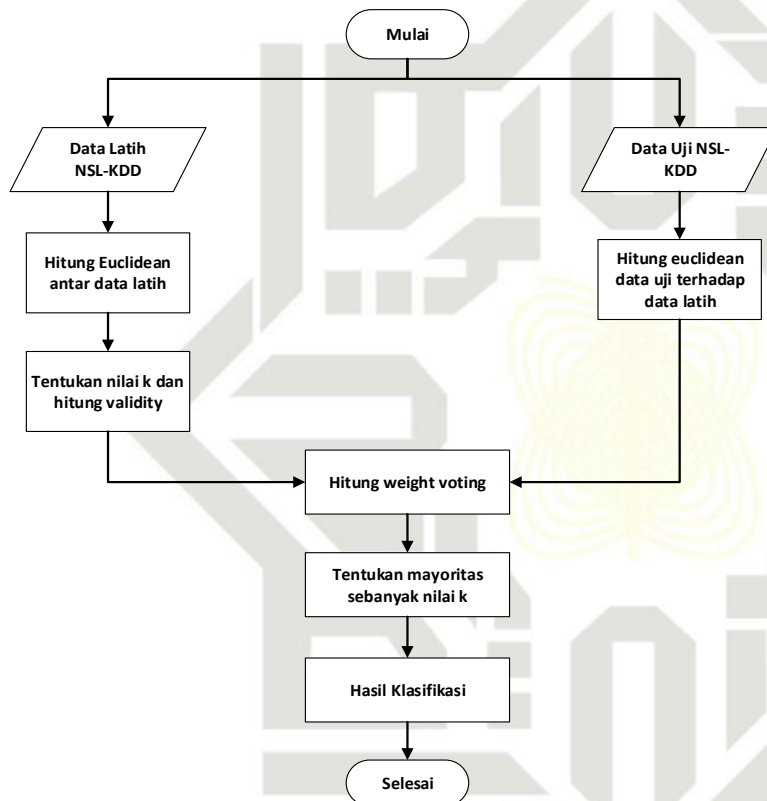
Data	A20	A21	A22	A23	A24	A25	A26	A27	A28	A29	Kelas
1	0	0	0.0980	0.17	0.0469	0.1848	0	0	0	0	1
2	0.0986	0	0.1020	0.1	0.0781	0	0	1	1	0	2
3	0	0	1	1	0	0.0326	0.1290	0.03	0.01	0.01	1
4	0	0.1268	1	1	0	0	0	0	0	0	1
5	0.0845	0	0.0353	0.04	0.0781	0	0	1	1	0	2
6	0.0845	0	0.0588	0.06	0.1094	0	0	1	1	0	2
7	0.0704	0	0.0902	0.09	0.0781	0	0	1	1	0	2
8	0.0845	0	0.0510	0.05	0.0938	0	0	1	1	0	2
9	0.0845	0	0.0510	0.05	0.1094	0	0	1	1	0	2
10	0	0.6056	0.8588	1	0	0.1304	0.0968	0	0	0	1
11	0.0845	0	0.0039	0	0.1094	0	0	1	1	0	2
12	0.0704	0	0.0078	0.01	0.0938	0	0	1	1	0	2
13	0	0.3099	1	1	0	0.0109	0.0645	0	0	0	1
14	0	0	1	1	0	0.0217	0.0968	0	0	0	1
15	0	0.2817	1	1	0	0.0109	0.1290	0	0	0	1
...
88.74	0	0	0.3020	0.3	0.0469	0.3261	0	0	0	0	1

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4.2.4 Data Mining

Pada tahapan ini ialah tahap perhitungan terhadap data NSL-KDD yang telah diproses pada tahap sebelumnya, perhitungan terhadap data NSL-KDD ini menggunakan metode MKNN (*Modified K-Nearest Neighbour*). Perhitungan menggunakan metode MKNN ini memiliki beberapa proses diantaranya ialah perhitungan jarak euclidean, menghitung validitas data latih dan menghitung nilai *weight voting* pada data. Berikut flowchart metode MKNN pada Gambar 4.5:



Gambar 4.5 Flowchart Klasifikasi MKNN

Berikut ialah detail keterangan *flowchart* klasifikasi serangan jaringan NSL-KDD menggunakan metode MKNN:

1. Data Latih

Data latih yang digunakan ialah data yang telah melewati beberapa tahapan proses KDD yaitu *selection*, *cleaning* dan *transformation*. Berikut diambil 10 sampel data latih yang digunakan untuk perhitungan manual pada Tabel 4.7:

Tabel 4.7 Data Latih NSL-KDD

Data	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19
1	0	0	0.0615	0	3.53×10^{-5}	0	0	0	0	0	0	0	0.0066	0.0079	0	0	0	0	1
2	0	0	0.0154	0.0909	0	0	0	0	0	0	0	0	0.4046	0.0238	1	1	0	0	0.05
3	0	0	0	0	1.67×10^{-5}	0.0016	0	1	0	0	0	0	0.0164	0.0198	0.2	0.2	0	0	1
4	0	0	0	0	1.43×10^{-5}	8.18×10^{-5}	0	1	0	0	0	0	0.0987	0.1270	0	0	0	0	1
5	0	0	0.0154	0.0909	0	0	0	0	0	0	0	0	0.5461	0.0357	1	1	0	0	0.05
6	0	0	0.0154	0.0909	0	0	0	0	0	0	0	0	0.3849	0.0635	1	1	0	0	0.14
7	0	0	0.8769	0.0909	0	0	0	0	0	0	0	0	0.8882	0.0913	1	1	0	0	0.09
8	0	0	0.0154	0.0909	0	0	0	0	0	0	0	0	0.4375	0.0317	1	1	0	0	0.06
9	0	0	0.0154	0.0909	0	0	0	0	0	0	0	0	0.6546	0.0119	1	1	0	0	0.02
10	0	0	0	0	2.06×10^{-5}	0.0004	0	1	0	0	0	0	0.0099	0.0278	0	0	0	0	1

Data	A20	A21	A22	A23	A24	A25	A26	A27	A28	A29	Kelas
1	0	0	0.0980	0.17	0.0469	0.1848	0	0	0	0	1
2	0.0986	0	0.1020	0.1	0.0781	0	0	1	1	0	2
3	0	0	1	1	0	0.0326	0.1290	0.03	0.01	0.01	1
4	0	0.1268	1	1	0	0	0	0	0	0	1
5	0.0845	0	0.0353	0.04	0.0781	0	0	1	1	0	2
6	0.0845	0	0.0588	0.06	0.1094	0	0	1	1	0	2
7	0.0704	0	0.0902	0.09	0.0781	0	0	1	1	0	2
8	0.0845	0	0.0510	0.05	0.0938	0	0	1	1	0	2
9	0.0845	0	0.0510	0.05	0.1094	0	0	1	1	0	2
10	0	0.6056	0.8588	1	0	0.1304	0.0968	0	0	0	1

2. Data Uji

Data uji yang digunakan ialah data yang telah disediakan oleh NSL-KDD yang juga telah melewati tahapan proses KDD diantaranya *selection*, *cleaning* dan *transformation*. Berikut diambil 10 sampel data uji yang digunakan untuk perhitungan manual pada

Tabel 4.8:

Tabel 4.8 Data Uji NSL-KDD

Data	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18
1	0	0	0	0	2.47×10^{-5}	0.0002	0	1	0	0	0	0	0.0296	0.0397	0	0	0	0
2	0	0	0.5846	0.0909	0	0	0	0	0	0	0	0	0.7336	0.0913	1	1	0	0
3	0	0	0.0154	0.0909	0	0	0	0	0	0	0	0	0.9211	0.0675	1	1	0	0
4	0	0	0	0	1.82×10^{-5}	0.0023	0	1	0	0	0	0	0.0263	0.0397	0	0	0	0

Hak Cipta Diindungi Undang-Undang

1. Diararag mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Diararag mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

0	0	0.5846	0.0909	0	0	0	0	0	0	0	0	0	0.8158	0.0079	1	1	0	0
0.0774	0	0.1231	0	3.14×10^{-5}	0.0028	0	1	0.0682	0	0	0	0	0.0033	0.0040	0	0	0	0
0	0	0.8769	0.0909	0	0	0	0	0	0	0	0	0	0.8882	0.0913	1	1	0	0
0	0	0.0154	0.0909	0	0	0	0	0	0	0	0	0	0.4375	0.0317	1	1	0	0
0	0	0	0	1.43×10^{-5}	0.0001	0	1	0	0	0	0	0	0.0987	0.1270	0	0	0	0
0	0	0	0	1.55×10^{-5}	0.0020	0	1	0	0	0	0	0	0.0461	0.0556	0	0	0	0

Data	A19	A20	A21	A22	A23	A24	A25	A26	A27	A28	A29	Class
1	1	0	0.28169	1	1	0	0.0109	0.1290	0	0	0	1
2	0.1	0.0704	0	0.0902	0.09	0.0781	0	0	1	1	0	2
3	0.06	0.0704	0	0.0667	0.07	0.0938	0	0	0.99	1	0	2
4	1	0	0.2817	1	1	0	0.0109	0.0645	0	0	0	1
5	0.01	0.0845	0	0.0078	0.01	0.0938	0	0	1	1	0	2
6	1	0	0	0.0980	0.1	0.0781	0	0	0.53	0	0.16	1
7	0.09	0.0704	0	0.0902	0.09	0.0781	0	0	1	1	0	2
8	0.06	0.0845	0	0.0510	0.05	0.0938	0	0	1	1	0	2
9	1	0	0.1268	1	1	0	0	0	0	0	0	1
10	1	0	0	1	1	0	0	0	0	0	0	1

c. Jarak data latih 2 dan 1

$$d(x,y) \text{ (data latih } x, \text{ data latih } y) = \sqrt{\sum_{i=1}^n (xi - yi)^2}$$

$$d(2,1) = \sqrt{(0-0)^2 + (0-0)^2 + (0,0154-0,0615)^2 + (0,0909-0)^2 + (0-3,53 \times 10^{-5})^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0,4046-0,0066)^2 + (0,0238-0,0079)^2 + (1-0)^2 + (1-0)^2 + (0-0)^2 + (0-0)^2 + (0,05-1)^2 + (0,0986-0)^2 + (0-0)^2 + (0,0120-0,0980)^2 + (0,1-0,17)^2 + (0,0781-0,0469)^2 + (0-0,1848)^2 + (0-0)^2 + (1-0)^2 + (1-0)^2 + (0-0)^2}$$

$$d(2,1) = 2,2630$$

Keseluruhan data latih dihitung dengan menggunakan perhitungan diatas hingga hingga memperoleh seluruh nilai jarak *euclidean* antar data latih. Berikut hasil perhitungan jarak *euclidean* antar data latih pada Tabel 4.9:

Tabel 4.9 Jarak *Euclidean* antar Data Latih

D	1	2	3	4	5	6	7	8	9	10	kelas
1	0	2.2630	1.6216	1.6066	2.2952	2.2260	2.5163	2.2672	2.3351	1.6289	1
2	2.2630	0	2.6292	2.7683	0.1685	0.1212	0.9916	0.0824	0.2642	2.8033	2
3	1.6216	2.6292	0	0.3647	2.6974	2.6254	2.8733	2.6660	2.7234	0.6917	1
4	1.6066	2.7683	0.3647	0	2.8286	2.7638	2.9854	2.8020	2.8512	0.5416	1
5	2.2952	0.1685	2.6974	2.8286	0	0.1918	0.9326	0.1118	0.1207	2.8644	2
6	2.2260	0.1212	2.6254	2.7638	0.1918	0	1.0009	0.1029	0.3000	2.7975	2
7	2.5163	0.9916	2.8733	2.9854	0.9326	1.0009	0	0.9764	0.9013	3.0338	2
8	2.2672	0.0824	2.6660	2.8020	0.1118	0.1029	0.9764	0	0.2222	2.8353	2
9	2.3351	0.2642	2.7234	2.8512	0.1207	0.3000	0.9013	0.2222	0	2.8900	2
10	1.6289	2.8033	0.6917	0.5416	2.8644	2.7975	3.0338	2.8353	2.8900	0	1



4. Perhitungan Validitas Data latih

Validitas data latih dihitung dengan menggunakan Persamaan (2.5) serta penggunaan nilai 0 dan 1 untuk fungsi S dengan menggunakan Persamaan (2.6). Langkah awal yang dilakukan untuk menghitung validitas data latih ialah menentukan nilai k yang digunakan sebagai acuan untuk menentukan jarak ketetanggaan antar data latih. Dalam hal ini nilai k yang digunakan ialah $k=3$. Selanjutnya bandingkan kelas pada label data (x) yang merupakan data acuan dan kelas pada label data ($Ni(x)$) sebagai data dengan jarak *Euclidean* terdekat terhadap data x . Kemudian apabila kedua kelas data bernilai sama maka nilai fungsi S ialah 1 dan apabila nilai kedua kelas berbeda maka nilai fungsi S nya ialah 0. Perbandingan dilakukan sebanyak nilai k yang ditentukan. Berikut ialah perhitungan validitas data latih yang mengacu data pada Tabel 4.9:

a. Validitas Data 1

$$Validity(data1) = \frac{1}{3} \sum_{i=1}^3 S((label\ S1\ (0), (label\ S4\ (1,6066), (label\ S3\ (1,6216))$$

$$Validity(data\ 1) = \frac{1}{3} * (1 + 1 + 1)$$

$$Validity(data\ 1) = 1$$



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

b. Validitas Data 2

$$Validity(data\ 2) = \frac{1}{3} \sum_{i=1}^3 S((label\ S2\ (0), (label\ S8\ (0,0824), (label\ S6\ (0,1212)))$$

$$Validity(data\ 2) = \frac{1}{3} * (1 + 1 + 1)$$

$$Validity(data\ 2) = 1$$

c. Validitas Data 3

$$Validity(data\ 3) = \frac{1}{3} \sum_{i=1}^3 S((label\ S3\ (0), (label\ S4\ (0,3647), (label\ S10\ (0,6917)))$$

$$Validity(data\ 3) = \frac{1}{3} * (1 + 1 + 1)$$

$$Validity(data\ 3) = 1$$

d. Validitas Data 4

$$Validity(data\ 4) = \frac{1}{3} \sum_{i=1}^3 S((label\ S4\ (0), (label\ S3\ (0,3647), (label\ S10\ (0,5416)))$$

$$Validity(data\ 4) = \frac{1}{3} * (1 + 1 + 1)$$

$$Validity(data\ 4) = 1$$

e. Validitas Data 5

$$Validity(data\ 5) = \frac{1}{3} \sum_{i=1}^3 S((label\ S5\ (0), (label\ S8\ (0,1119), (label\ S9\ (0,1207)))$$

$$Validity(data\ 5) = \frac{1}{3} * (1 + 1 + 1)$$

$$Validity(data\ 5) = 1$$

f. Validitas Data 6

$$Validity(data\ 6) = \frac{1}{3} \sum_{i=1}^3 S((label\ S6\ (0), (label\ S8\ (0,1029), (label\ S2\ (0,1212)))$$

$$Validity(data\ 6) = \frac{1}{3} * (1 + 1 + 1)$$

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$Validity(data\ 6) = 1$$

g. Validitas Data 7

$$Validity(data\ 7) = \frac{1}{3} \sum_{i=1}^3 S((label\ S7\ (0), (label\ S9\ (0,9013), (label\ S5\ (0,9326))$$

$$Validity(data\ 7) = \frac{1}{3} * (1 + 1 + 1)$$

$$Validity(data\ 7) = 1$$

h. Validitas Data 8

$$Validity(data\ 8) = \frac{1}{3} \sum_{i=1}^3 S((label\ S8\ (0), (label\ S2\ (0,0824), (label\ S6\ (0,1029))$$

$$Validity(data\ 8) = \frac{1}{3} * (1 + 1 + 1)$$

$$Validity(data\ 8) = 1$$

i. Validitas Data 9

$$Validity(data\ 9) = \frac{1}{3} \sum_{i=1}^3 S((label\ S9\ (0), (label\ S5\ (0,1207), (label\ S8\ (0,2222))$$

$$Validity(data\ 9) = \frac{1}{3} * (1 + 1 + 1)$$

$$Validity(data\ 9) = 1$$

j. Validitas Data 10

$$Validity(data\ 10) = \frac{1}{3} \sum_{i=1}^3 S((label\ S10\ (0), (label\ S4\ (0,5416), (label\ S3\ (0,6917))$$

$$Validity(data\ 10) = \frac{1}{3} * (1 + 1 + 1)$$

$$Validity(data\ 10) = 1$$

Proses perhitungan di atas dilakukan pada seluruh data latih hingga memperoleh nilai validitas keseluruhan data latih NSL KDD. Berikut ialah hasil perhitungan validitas data latih pada Tabel 4.10

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tabel 4.10 Nilai Validitas Data Latih

Data	K1	K2	K3	Validitas
1	1	1	1	1
2	1	1	1	1
3	1	1	1	1
4	1	1	1	1
5	1	1	1	1
6	1	1	1	1
7	1	1	1	1
8	1	1	1	1
9	1	1	1	1
10	1	1	1	1

5. Perhitungan Jarak *Euclidean* Data Uji terhadap Data Latih

Perhitungan jarak *Euclidean* data uji terhadap data latih dilakukan dengan menggunakan persamaan yang sama dengan perhitungan antar data latih yakni persamaan 2.3. Berikut perhitungan jarak euclidean antara data uji (Tabel 4.8) dan data latih (Tabel 4.7):

a. Jarak *Euclidean* Data Uji 1 dan Data Latih 1

$$d(1,1) = \sqrt{(0-0)^2 + (0-0)^2 + (0-0,0615)^2 + (0-0)^2 + (2,47 \times 10^{-5} - 3,53 \times 10^{-5})^2 + (0,0002-0)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0,0296-0,0066)^2 + (0,0397-0,0079)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (1-1)^2 + (0-0)^2 + (0,2817-0)^2 + (1-0,0980)^2 + (1-0,17)^2 + (0-0,0469)^2 + (0,0109-0,1848)^2 + (0,1290-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2}$$

$$d(1,1) = 1,6236$$

b. Jarak *Euclidean* Data Uji 1 dan Data Latih 2

$$d(1,2) = \sqrt{(0-0)^2 + (0-0)^2 + (0-0,0154)^2 + (0-0,0909)^2 + (2,47 \times 10^{-5} - 0)^2 + (0,0002 - 0)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0,0296 - 0,4046)^2 + (0,0397 - 0,0238)^2 + (0-1)^2 + (0-1)^2 + (0-0)^2 + (0-0)^2 + (1-0,05)^2 + (0-0,0986)^2 + (0,2817 - 0)^2 + (1-0,1020)^2 + (1-0,1)^2 + (0-0,0781)^2 + (0,0109 - 0)^2 + (0,1290 - 0)^2 + (0-1)^2 + (0-1)^2 + (0-0)^2}$$

$$d(1,2) = 2,7893$$

c. Jarak *Euclidean* Data Uji 2 dan Data Latih 4

$$d(2,4) = \sqrt{(0-0)^2 + (0-0)^2 + (0,5846 - 0)^2 + (0,0909 - 0)^2 + (0 - 1,43 \times 10^{-5})^2 + (0 - 8,18 \times 10^{-5})^2 + (0-0)^2 + (0-1)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0,7336 - 0,0987)^2 + (0,0913 - 0,1270)^2 + (1-0)^2 + (1-0)^2 + (0-0)^2 + (0-0)^2 + (0,1-1)^2 + (0,0704 - 0)^2 + (0-0,1268)^2 + (0,0902 - 1)^2 + (0,09 - 1)^2 + (0,0781 - 0)^2 + (0-0)^2 + (0-0)^2 + (1-0)^2 + (1-0)^2 + (0-0)^2}$$

$$d(2,4) = 2,8718$$

Perhitungan diatas diterapkan terhadap keseluruhan antar data uji dan data latih. Berikut ialah hasil perhitungan jarak *euclidean* antara data uji dan data latih pada Tabel 4.11:

Tabel 4.11 Jarak *Euclidean* Data Uji dan Data Latih

No	de 1	de 2	de 3	de 4	de 5	de 6	de 7	de 8	de 9	de 10
1	1.6236	2.3820	2.4013	1.6197	2.4483	1.1556	2.5163	2.2672	1.6066	1.5957
2	2.7893	0.6636	0.5215	2.7875	0.7158	2.3076	0.9916	0.0824	2.7683	2.7700
3	0.4017	2.7505	2.7680	0.4068	2.8580	1.4155	2.8733	2.6660	0.3647	0.3176
4	0.2306	2.8718	2.8841	0.2028	2.9757	1.4059	2.9854	2.8020	0	0.1547
5	2.8530	0.6087	0.3797	2.8514	0.6333	2.3376	0.9326	0.1118	2.8286	2.8332
6	2.7856	0.6716	0.5428	2.7838	0.7315	2.2697	1.0009	0.1029	2.7638	2.7661
7	3.0180	0.3308	0.8638	3.0168	0.3430	2.5361	0	0.9764	2.9854	2.9971

8	2.8238	0.6484	0.4858	2.8220	0.6882	2.3101	0.9764	0	2.8020	2.8045
9	2.8773	0.5893	0.2773	2.8758	0.5948	2.3767	0.9013	0.2222	2.8512	2.8572
10	0.3752	2.9173	2.9331	0.3750	3.0155	1.4475	3.0338	2.8353	0.5416	0.6443

6. Perhitungan *Weight Voting*

Proses perhitungan *weight voting* dilakukan dengan menggunakan Persamaan (2.9) dengan menghitung nilai validitas data latih (Tabel 4.10) dan hasil perhitungan jarak *euclidean* data uji terhadap data latih (Tabel 4.11). Hasil perhitungan *weight voting* ini nantinya dijadikan sebagai acuan untuk penentuan klasifikasi data dengan mengambil nilai tertinggi sebanyak nilai k. Berikut ialah langkah-langkah perhitungan *weight voting*:

$$W(\text{data uji 1,1}) = 1 * \frac{1}{1,6236 + 0,5}$$

$$= 0,4709$$

$$W(\text{data uji 1,2}) = 1 * \frac{1}{2,7893 + 0,5}$$

$$= 0,3040$$

$$W(\text{data uji 2,1}) = 1 * \frac{1}{2,3820 + 0,5}$$

$$= 0,3740$$

$$W(\text{data uji 2,2}) = 1 * \frac{1}{0,6636 + 0,5}$$

$$= 0,8594$$

$$W(\text{data uji 3,1}) = 1 * \frac{1}{2.4013 + 0,5}$$

$$= 0,3447$$

Langkah perhitungan diatas diterapkan pada seluruh data jarak *euclidean* data uji dan data latih pada Tabel 4.10 hingga diperoleh keseluruhan nilai *weight voting*. Berikut ialah hasil perhitungan *weight voting* pada Tabel 4.12:

Tabel 4.12 Nilai Weight Voting

No	WV 1	WV 2	WV 3	WV 4	WV 5	WV 6	WV 7	WV 8	WV 9	WV 10
1	0.4709	0.347	0.3447	0.4718	0.3392	0.6040	0.3315	0.3614	0.4747	0.4772
2	0.3040	0.8594	0.9789	0.3042	0.8225	0.3562	0.6704	1.7171	0.3060	0.3058
3	1.1090	0.3076	0.3060	1.1028	0.2978	0.5221	0.2964	0.3159	1.1564	1.2231
4	1.3688	0.2966	0.2955	1.4228	0.2877	0.5247	0.2869	0.3028	2	1.5273
5	0.2982	0.9020	1.1368	0.2984	0.8824	0.3524	0.6980	1.6346	0.3004	0.300
6	0.3044	0.8535	0.9590	0.3045	0.8120	0.3611	0.6662	1.6587	0.3064	0.3062
7	0.2843	1.2036	0.7333	0.2843	1.1862	0.3294	2	0.6773	0.2869	0.2860
8	0.3009	0.8708	1.0144	0.3010	0.8416	0.3559	0.6773	2	0.3028	0.3026
9	0.2961	0.9180	1.2866	0.2962	0.9134	0.3476	0.7136	1.3847	0.2984	0.2979
10	1.1427	0.2926	0.2913	1.1429	0.2845	0.5135	0.2830	0.2998	0.9600	0.8739

7. Penentuan Mayoritas Data

Pada tahapan ini data pada Tabel 4.12 diurutkan dari nilai tertinggi ke nilai terkecil kemudian data diambil sebanyak nilai *k*. Dalam hal ini nilai *k* yang digunakan ialah *k* = 3, sehingga 3 nilai tertinggi menjadi acuan untuk penentuan kelas hasil klasifikasi. Berikut ialah data nilai *weight voting* yang telah diurutkan dari nilai tertinggi ke terkecil serta kelas dari masing-masing data yang dapat dilihat pada Tabel 4.13:

Tabel 4.13 Mayoritas Data

No	WV 1	Kelas	WV 2	Kelas	WV 3	Kelas	WV 4	Kelas	WV 5	Kelas	WV 6	Kelas	WV 7	Kelas	WV 8	Kelas	WV 9	Kelas	WV 10	Kelas
1	1.3688	1	1.2036	2	1.2866	1	1.4228	1	1.1862	2	0.604	1	2	2	2	2	2	1	1.5273	1
2	1.1427	1	0.9180	1	1.1368	2	1.1429	1	0.9134	1	0.5247	1	0.7136	1	1.7171	2	1.1564	2	1.2231	2
3	1.1090	2	0.9020	2	1.0144	2	1.1028	2	0.8824	2	0.5221	2	0.6980	2	1.6587	1	0.9600	1	0.8739	1
4	0.4709	1	0.8708	2	0.9789	2	0.4718	1	0.8416	2	0.5135	1	0.6773	2	1.6346	2	0.4747	1	0.4772	1
5	0.3044	1	0.8594	2	0.9590	1	0.3045	1	0.8225	2	0.3611	1	0.6704	2	1.3847	1	0.3064	1	0.3062	1
6	0.304	2	0.8535	1	0.7333	2	0.3042	2	0.8120	1	0.3562	2	0.6662	1	0.6773	2	0.3060	2	0.3058	2
7	0.3009	2	0.3470	1	0.3447	1	0.3010	2	0.3392	1	0.3559	2	0.3315	1	0.3614	1	0.3028	2	0.3026	2
8	0.2982	2	0.3076	2	0.306	2	0.2984	2	0.2978	2	0.3524	2	0.2964	2	0.3159	2	0.3004	2	0.300	2
9	0.2961	1	0.2966	1	0.2955	1	0.2962	1	0.2877	1	0.3476	1	0.2869	1	0.3028	1	0.2984	1	0.2979	1
10	0.2843	2	0.2926	1	0.2913	1	0.2843	2	0.2845	1	0.3294	2	0.283	1	0.2998	1	0.2869	2	0.286	2

Berdasarkan Tabel 4.13 diatas diambil 3 data nilai tertinggi dan tentukan kelas yang merupakan mayoritas dari data. Berikut ialah hasil klasifikasi dengan menentukan mayoritas data sebanyak nilai $k=3$ pada Tabel 4.14:

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tabel 4.14 Hasil Klasifikasi

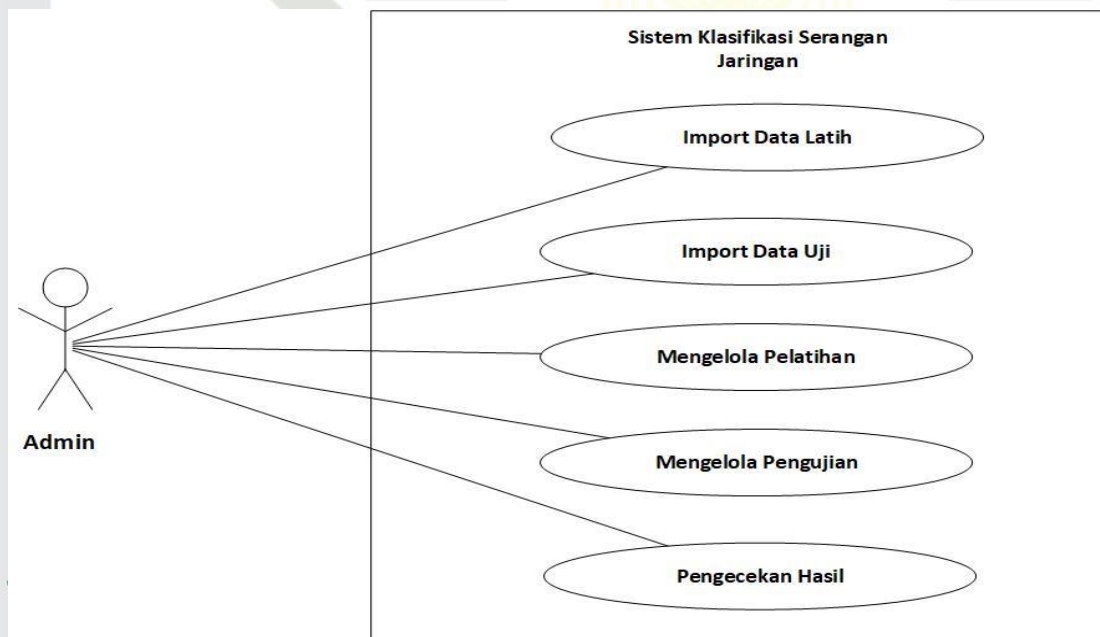
K	WV 1	WV 2	WV 3	WV 4	WV 5	WV 6	WV 7	WV 8	WV 9	WV 10
K1	1	2	1	1	2	1	2	2	1	1
K2	1	1	2	1	1	1	1	2	2	2
K3	2	2	2	2	2	2	2	1	1	1
Mayoritas	1	2	2	1	2	1	2	2	1	1

4.3 Analisa Sistem

Analisa sistem pada penelitian ini menggunakan pemodelan berorientasi objek yaitu UML (*Unified Modelling Language*). Tahapan-tahapan analisa pada UML ini sendiri ialah: *use case diagram*, *sequence diagram*, *class diagram* dan *activity diagram*.

4.3.1 Use Case Diagram

Use Case Diagram menggambarkan proses interaksi apa saja yang terjadi antara pengguna terhadap sistem yang akan dibangun pada penelitian ini. Berikut *use case diagram* pada Gambar 4.6:



Gambar 4.6 Use Case Diagram Sistem Klasifikasi Serangan Jaringan

Berdasarkan Gambar 4.6 diatas analisa *use case diagram* terdiri dari satu aktor yakni Admin dan 5 *use case* yaitu proses *import* data latih, *import* data

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

uji, pelatihan, pengujian dan pengecekan hasil klasifikasi sistem. Proses keseluruhan *use case* tersebut dijabarkan secara rinci pada *use case* spesifikasi berikut:

Use case Spesifikasi *Import* Data Latih

Import data latih merupakan langkah awal yang dilakukan oleh *user* sebelum dilakukan nya proses pelatihan. Berikut *use case* spesifikasi *import data latih* pada Tabel 4.15:

Tabel 4.15 Use Case Spesifikasi *Import* data latih

<i>Use case</i>	<i>Import</i> data latih
Aktor	Admin
Kondisi Awal	Data latih belum ada
Kondisi Akhir	Data latih berhasil disimpan
Main Success Scenario	<ol style="list-style-type: none"> 1. <i>Use case</i> dimulai saat <i>user</i> akan melakukan <i>import</i> data latih. 2. <i>User</i> menekan tombol <i>upload file</i> pada sistem 3. Sistem menampilkan <i>form upload</i> untuk memasukkan data latih 4. <i>User</i> melakukan <i>upload</i> data latih ke sistem dalam bentuk file csv dan menekan tombol <i>upload file</i> 5. Sistem menyimpan data latih yang telah di <i>upload</i> oleh <i>user</i> ke <i>database</i> 6. Data latih berhasil disimpan pada sistem
Alternative Scenario	Admin batal mengupload file data latih
Exception	Gagal! Jenis file tidak dapat ditermima. Admin salah menginputkan format data latih.

Use Case Spesifikasi *Import* Data Uji

Use case spesifikasi berikut mendeskripsikan proses *import* data uji ke dalam sistem. Berikut *use case* spesifikasi *import data uji* pada Tabel 4.16:

Tabel 4.16 Use Case Spesifikasi *Import* data uji

<i>Use case</i>	<i>Import</i> data uji
Aktor	Admin
Kondisi Awal	Data uji belum ada
Kondisi Akhir	Data uji berhasil disimpan
Main Success Scenario	<ol style="list-style-type: none"> 1. <i>Use case</i> dimulai saat <i>user</i> akan melakukan <i>import</i> data uji. 2. <i>User</i> menekan tombol <i>uoload file</i> pada sitem 3. Sistem menampilkan <i>form upload</i> untuk menginputkan data uji 4. <i>User</i> melakukan <i>upload file</i> data uji ke dalam sistem dalam bentuk csv dan menekam tombol <i>upload file</i> 5. Sistem menyimpan data uji yang telah di <i>upload</i> oleh <i>user</i> ke <i>database</i>

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

	6. Data uji berhasil disimpan pada sistem
<i>Alternative Scenario</i>	Admin batal mengupload file data uji
<i>Exception</i>	Gagal! Jenis file tidak dapat diterima. Admin salah menginputkan format data uji.

3. Use Case Spesifikasi Pelatihan

Use case spesifikasi pelatihan ialah proses pelatihan data latih setelah data latih di *import* ke dalam *database* sistem. Proses pelatihan dilakukan dengan menghitung jarak ketetanggaan antar data latih sesuai dengan nilai *k* yang ditentukan oleh *user*. Berikut *use case* spesifikasi pelatihan pada Tabel 4.17:

Tabel 4.17 Use Case Spesifikasi Pelatihan

<i>Use case</i>	Mengeola Pelatihan
Aktor	Admin
Kondisi Awal	Data latih telah terdapat pada <i>database</i> sistem
Kondisi Akhir	Data berhasil dilakukan pelatihan
<i>Main Success Scenario</i>	<ol style="list-style-type: none"> 1. <i>Use case</i> dimulai ketika <i>user</i> akan melakukan proses pelatihan 2. <i>User</i> memilih menu pelathan kemudian sistem menampilkan halaman pelatihan 3. <i>User</i> memilih inputan nilai <i>k</i> dengan jumlah data latih yang telah ditentukan 4. <i>User</i> menekan tombol mulai pelathan 5. Sistem melakukan pelatihan 6. Sistem berhasil melakukan pelatihan dan menampilkan jumlah data latih yang telah berhasil dihitung validitas nya.

4. Use Case Spesifikasi Pengujian

Use case spesifikasi pengujian merupakan proses pengujian data setelah nilai validitas data latih didapatkan. Pengujian dilakukan dengan menghitung jarak ketetanggaan antar data uji dan data latih sesuai dengan nilai *k* yang telah ditentukan *user*. Berikut ialah *Use case* spesifikasi pengujian pada Tabel 4.18:

Tabel 4.18 Use Case Spesifikasi Pengujian

<i>Use case</i>	Mengelola Pengujian
Aktor	Admin
Kondisi Awal	Data belum diuji dan telah terdapat pada <i>database</i> sistem
Kondisi Akhir	Data berhasil dilakukan pengujian
<i>Main Success Scenario</i>	<ol style="list-style-type: none"> 1. <i>Use case</i> dimulai saat <i>user</i> akan melakukan proses pengujian 2. <i>User</i> memilih menu pengujian dan sistem menampilkan halaman pengujian

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

	<ol style="list-style-type: none"> 3. <i>User</i> memilih inputan nilai k dengan jumlah data uji yang telah ditentukan 4. <i>User</i> menekan tombol mulau pengujian 5. Sistem melakukan pengujian 6. Sistem berhasil melakukan pengujian serta menampilkan persentase akurasi pengujian, <i>confusion matrix</i> dan jumlah data yang telah diujikan
--	---

5. *Use Case* Spesifikasi Pengecekan Hasil

Use case ini berisi hasil pengujian yang telah dilakukan pada tahap pengujian. Informasi yang ditampilkan pada *use case* ini ialah berupa grafik dan *confusion matrix* serta detail dari masing-masing nilai k hasil pengujian. Berikut *use case* spesifikasi pengecekan hasil pada Tabel 4.19:

Tabel 4.19 *Use Case* Spesifikasi Hasil

<i>Use case</i>	Pengecekan Hasil
Aktor	Admin
Kondisi Awal	Data telah diklasifikasikan
Kondisi Akhir	<i>User</i> dapat melihat detail hasil pengujian
<i>Main Success Scenario</i>	<ol style="list-style-type: none"> 1. <i>Use case</i> dimulai ketika <i>user</i> akan melakukan pengecekan hasil pengujian 2. <i>User</i> memilih menu riwayat pengujian 3. Sistem menampilkan halaman hasil pengujian berupa grafik, informasi waktu pengujian, jumlah data latih dan uji yang digunakan, nilai k, akurasi serta detail hasil pengujian dari masing-masing nilai k

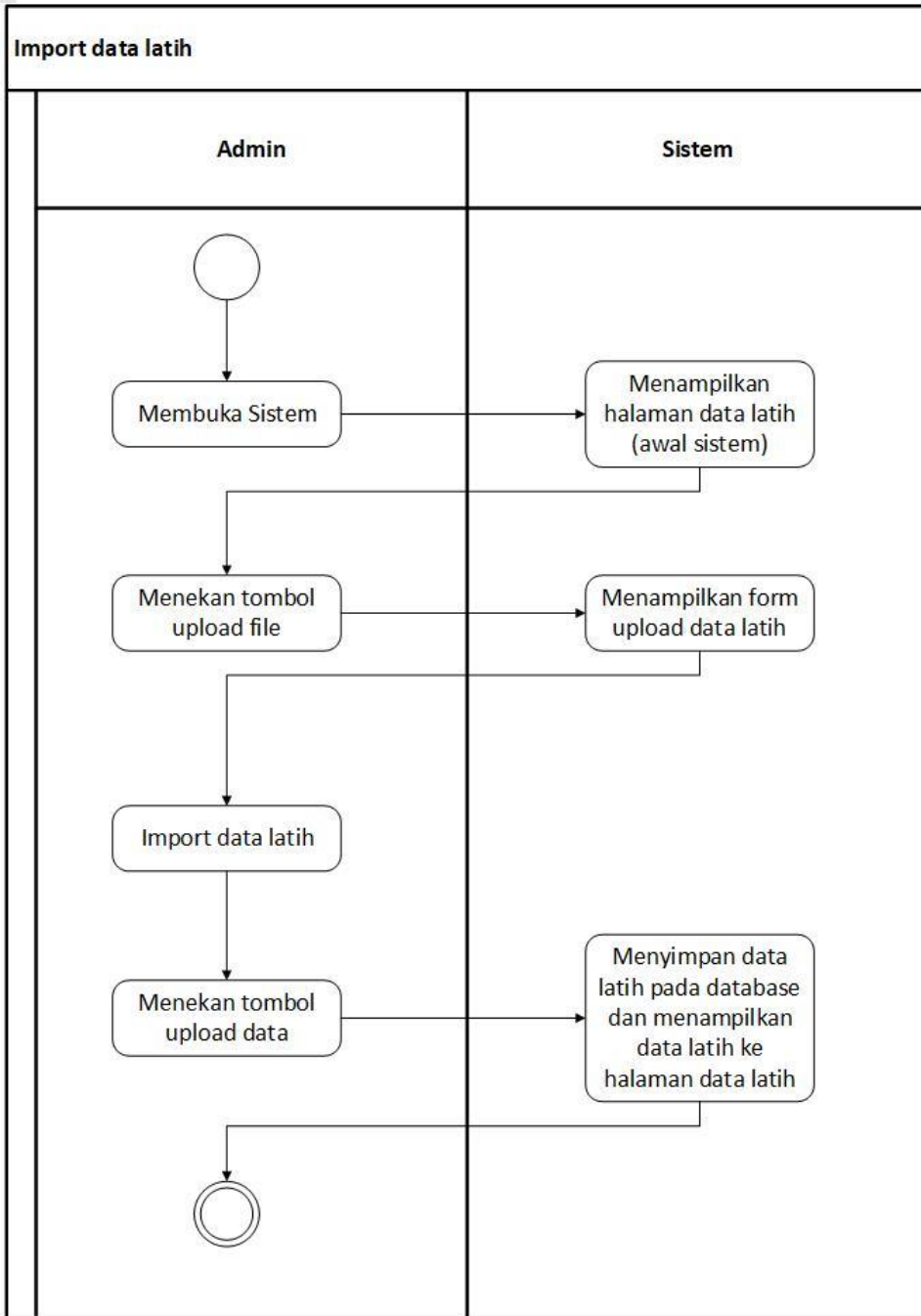
4.3.2 Activity Diagram

Activity diagram ialah gambaran alur kerja sistem yang divisualisasikan dalam bentuk *workflow*. *Activity diagram* dibuat berdasarkan *use case diagram* yang merupakan awal analisa sistem pada pemodelan UML. Berikut ialah *activity diagram* terhadap sistem yang akan dibangun pada penelitian ini:

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1. Activity Diagram Import Data Latih

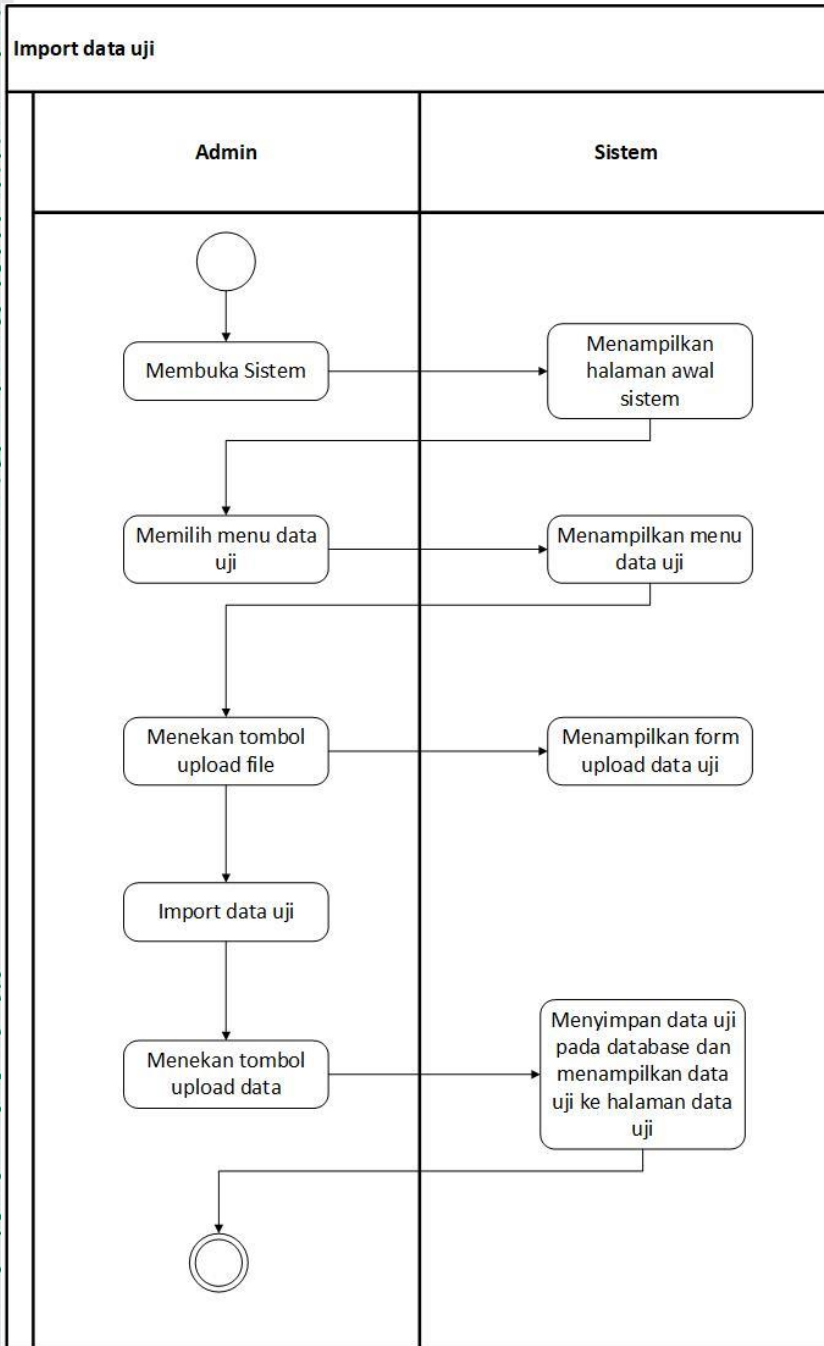


Gambar 4.7 Activity Diagram Import Data Latih

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2. Activity Diagram Import Data Uji

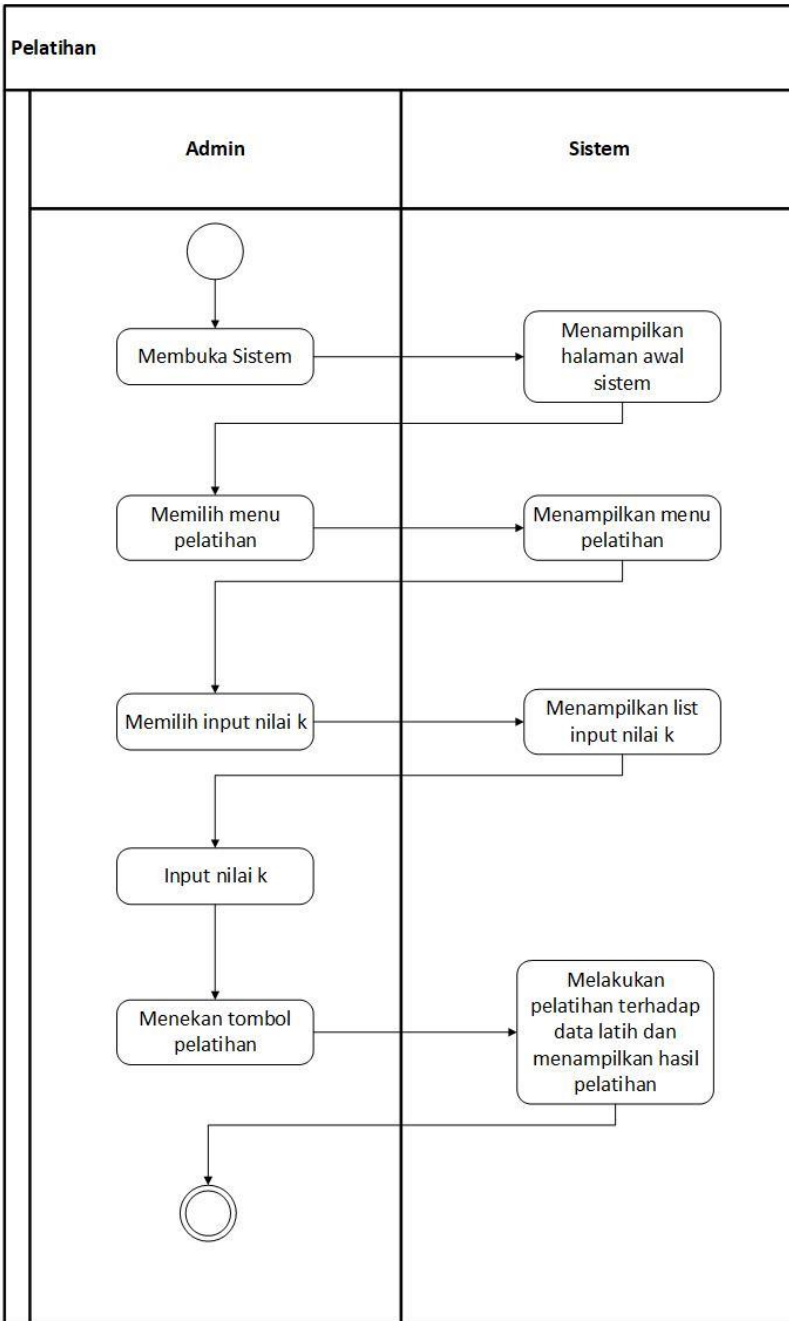


Gambar 4.8 Activity Diagram Import Data Uji

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3. Activity Diagram Pelatihan

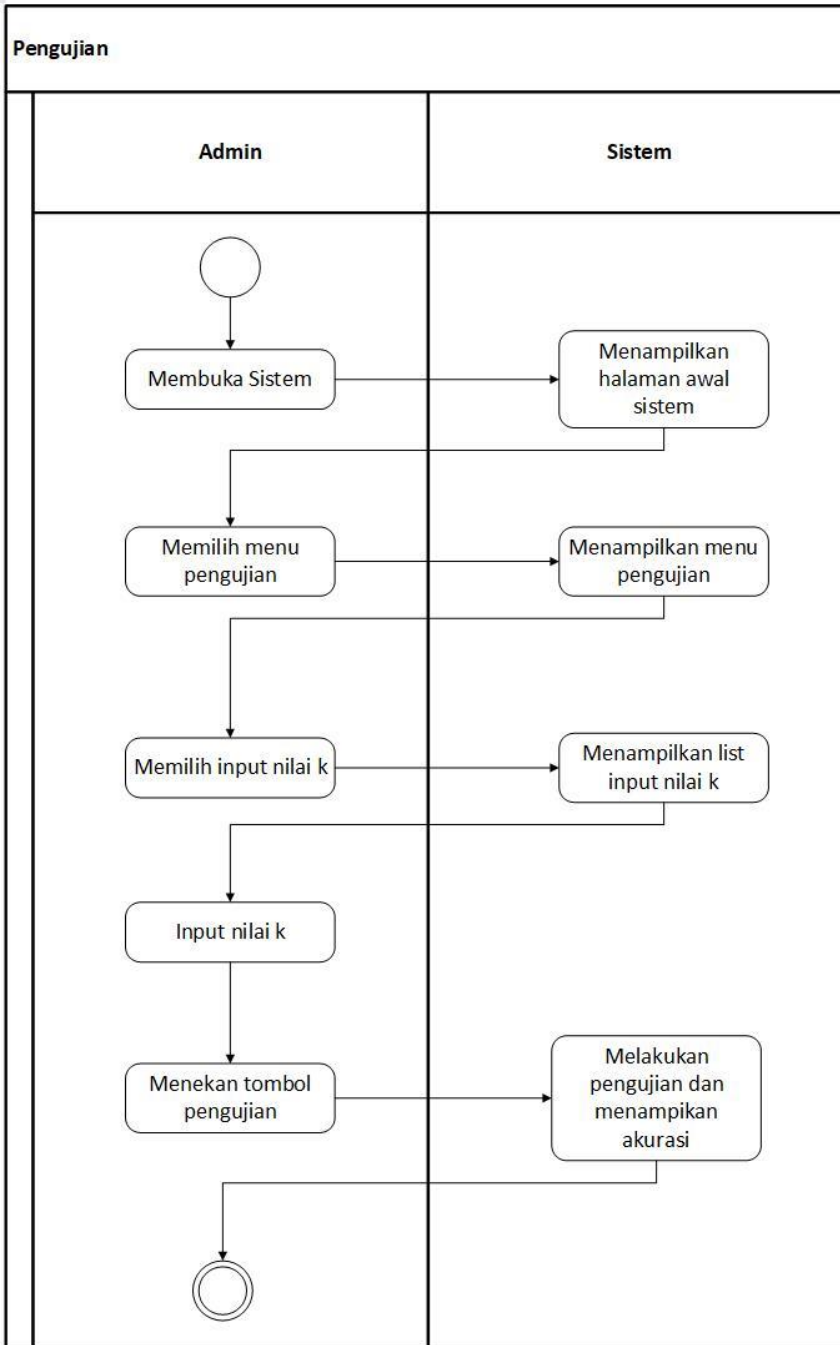


Gambar 4.9 Activity Diagram Pelatihan

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4. Activity Diagram Pengujian

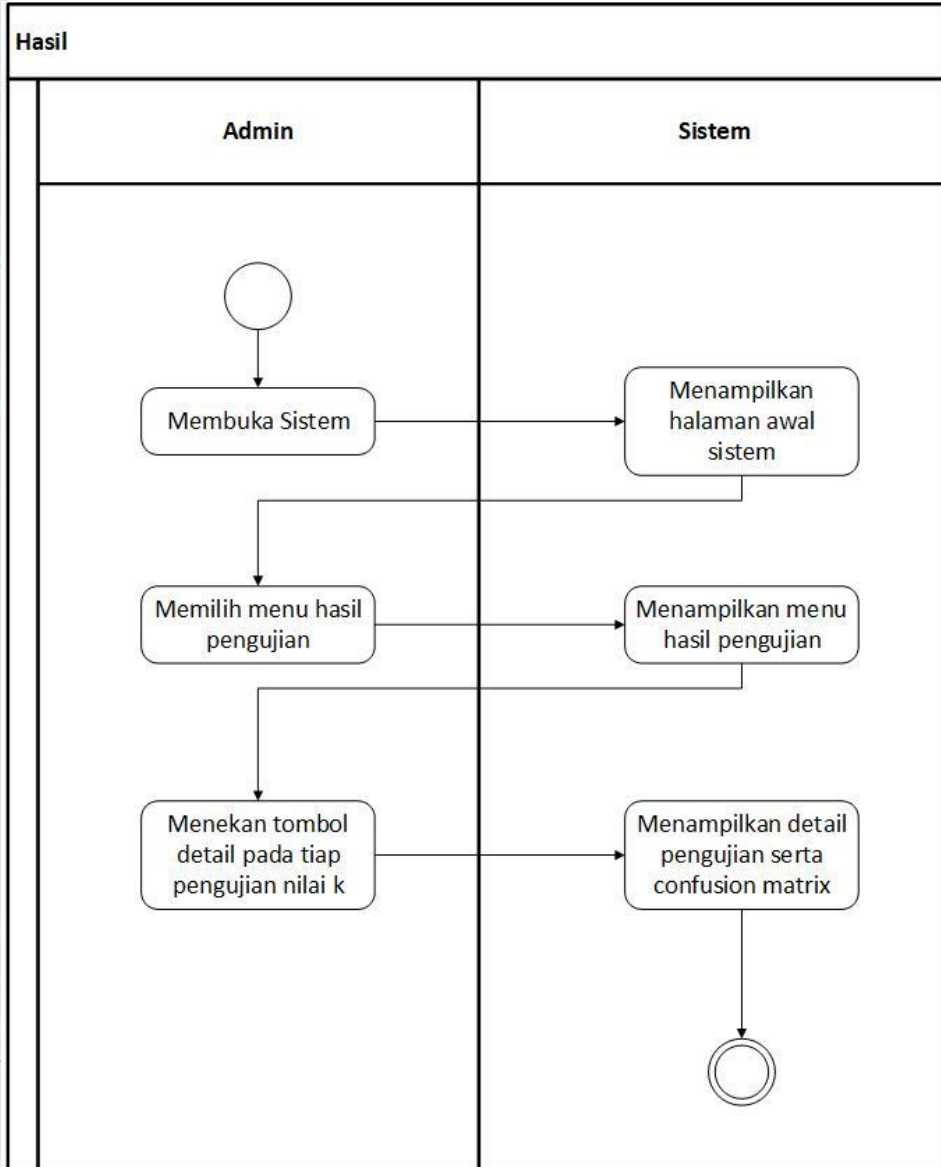


Gambar 4.10 Activity Diagram Pengujian

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Activity Diagram Pengecekan Hasil



Gambar 4.11 Activity Diagram Pengecekan Hasil

4.3.1 Sequence Diagram

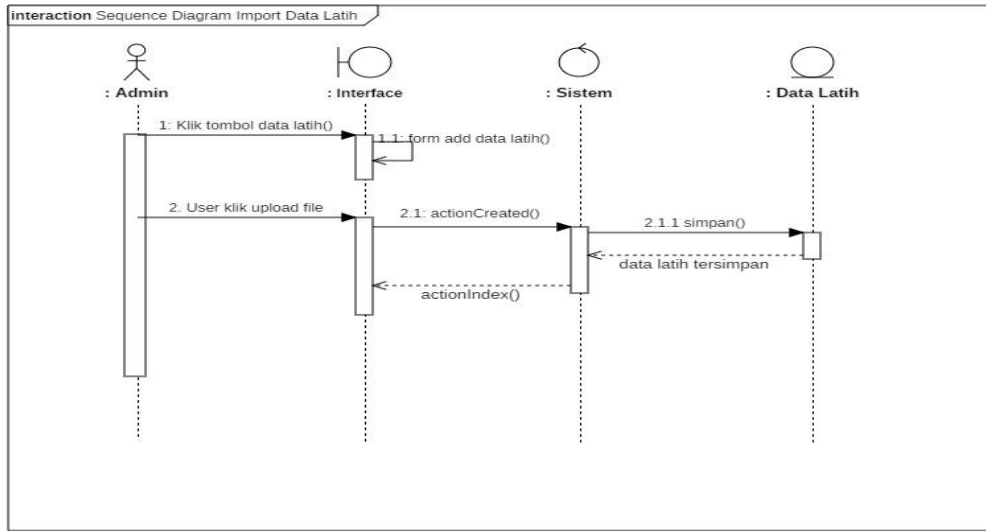
Sequence diagram menggambarkan hubungan atau interaksi antara *user* dan sistem terhadap urutan alur proses sistem serta gambaran antara objek pada sistem yang akan dibangun.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1. Sequence Diagram Import Data Latih

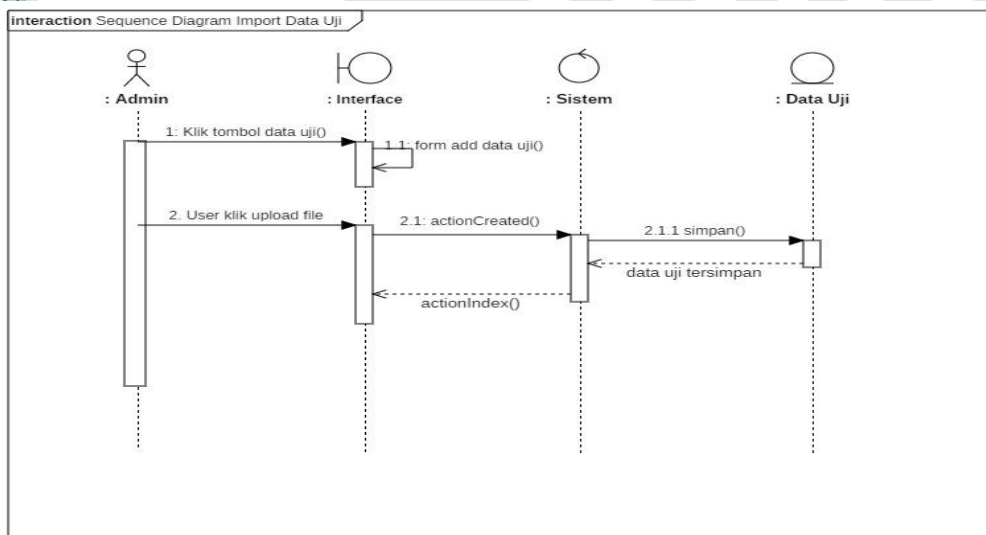
Sequence Diagram Import data latihan berikut mendeskripsikan tahapan yang dilakukan oleh user dalam melakukan import data latihan ke dalam sistem. Sequence diagram import data latihan dapat dilihat pada Gambar 4.12:



Gambar 4.12 Sequence Diagram Import Data Latih

2. Sequence Diagram Import Data Uji

Sequence Diagram Import data uji berikut mendeskripsikan tahapan yang dilakukan oleh user dalam melakukan import data uji ke dalam sistem. Sequence diagram import data uji dapat dilihat pada Gambar 4.13:



Gambar 4.13 Sequence Diagram Import Data Uji

Hak Cipta Dilindungi Undang-Undang

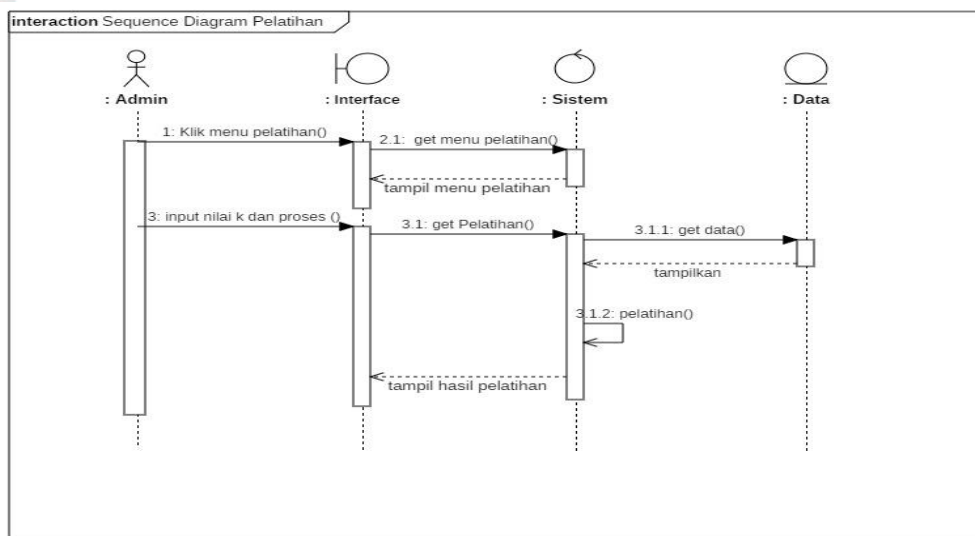
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3. Sequence Diagram Pelatihan

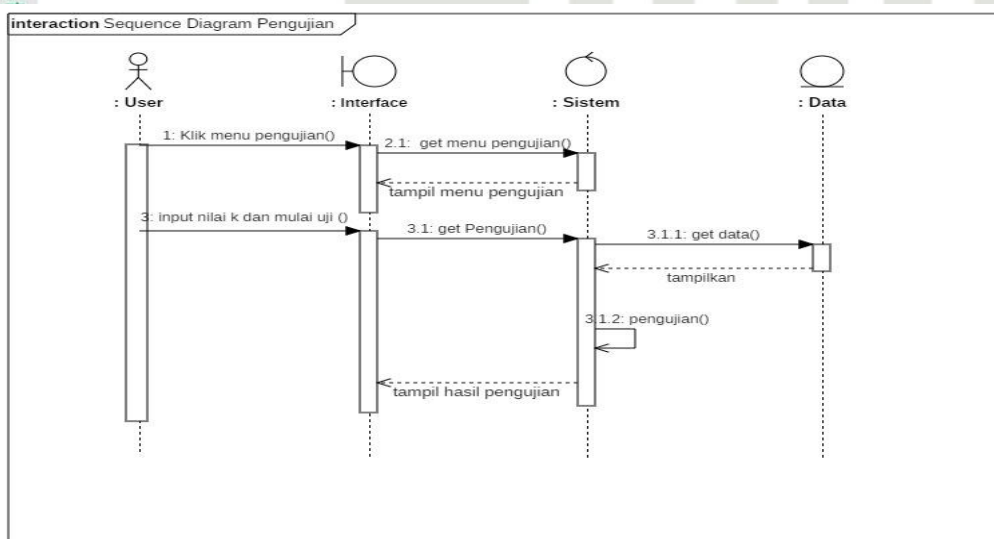
Sequence diagram pelatihan menggambarkan tahapan urutan proses ketika user melakukan pelatihan pada sistem. Berikut sequence diagram pelatihan pada gambar 4.14:



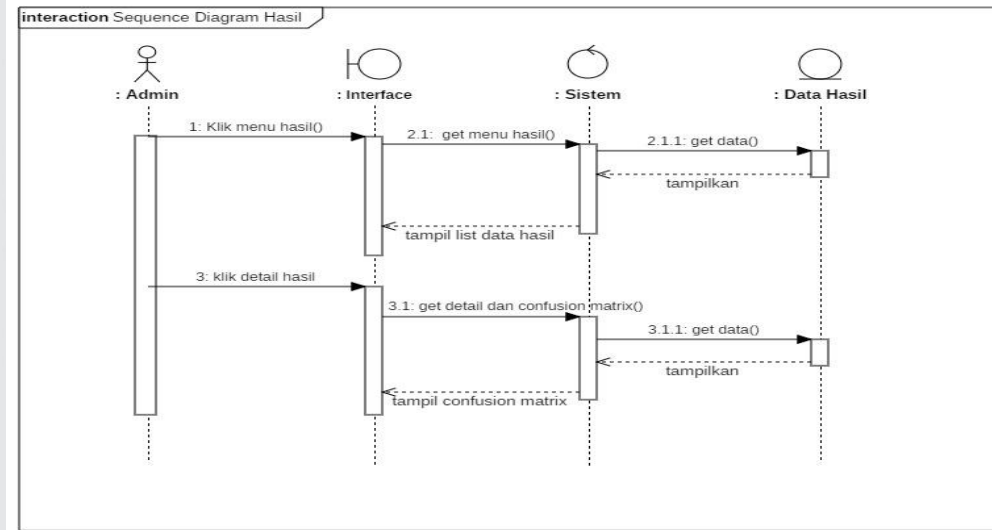
Gambar 4.14 Sequence Diagram Pelatihan

4. Sequence Diagram Pengujian

Sequence diagram pengujian menggambarkan tahapan urutan proses ketika user melakukan pengujian pada sistem. Berikut sequence diagram pengujian pada gambar 4.15:



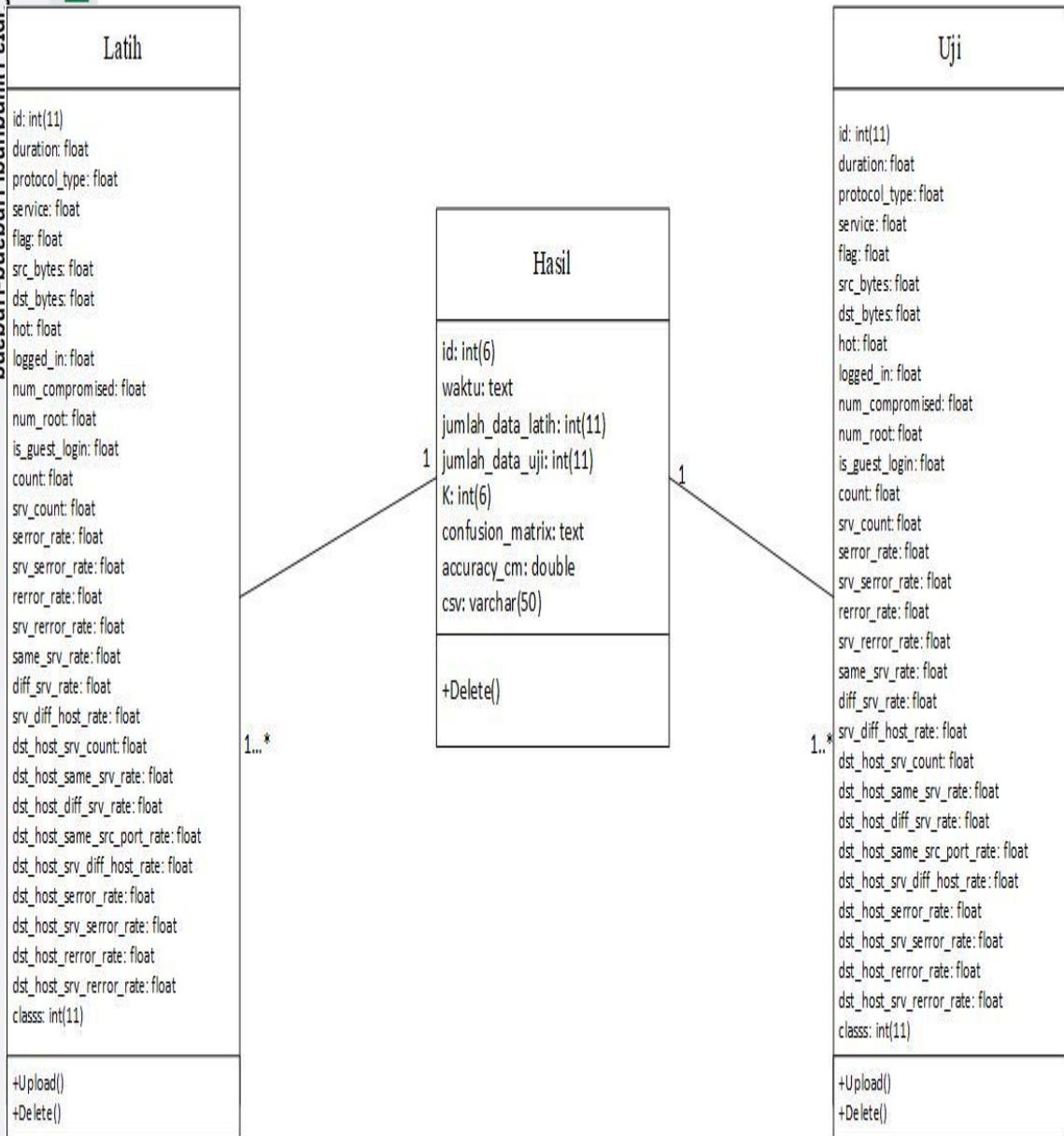
Gambar 4.15 Sequence Diagram Pengujian



Gambar 4.16 Sequence Diagram Pengecekan Hasil

4.3.2 Class Diagram

Class diagram memetakan struktur sistem dan class serta menggambarkan hubungan antara class sistem yang akan dibangun. Berikut ialah class diagram pada sistem yang akan dibangun pada Gambar 4.17



Gambar 4.17 Class Diagram Sistem Klasifikasi Serangan

Berdasarkan Gambar 4.17 diatas sistem yang akan dibangun memiliki 3 kelas yakni:

Class latih ialah *class* yang digunakan untuk melakukan pengolahan data latih pada sistem yang akan dibangun.

Class uji merupakan *class* yang digunakan untuk melakukan pengolahan data uji terhadap sistem yang akan dibangun.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3. *Class* hasil ialah *class* yang digunakan untuk menunjukkan hasil klasifikasi setelah proses pengujian dilakukan pada sistem yang akan dibangun.

4.4 Perancangan Sistem

Pada tahap ini dijabarkan mengenai perancangan sistem klasifikasi serangan jaringan dengan menerapkan metode MKNN yang akan dibangun. Perancangan yang akan dijabarkan meliputi perancangan *database*, perancangan struktur menu dan perancangan *interface* atau antar muka sistem.

4.4.1 Perancangan Database

Perancangan tabel *database* sistem mengacu kepada *class* diagram pada gambar 4.17, sehingga perancangan tabel yang dibentuk ialah:

1. Nama : Latih
- Deskripsi : Berisi data latih NSL-KDD
- Primary Key* : id

Berikut ialah detail tabel perancangan *database* data latih pada Tabel 4.20 berikut:

Tabel 4.20 Tabel Data Latih

Atribut	Type dan Length	Keterangan
id	int(11)	<i>Primary key</i>
<i>duration</i>	<i>float</i>	
<i>protocol_type</i>	<i>float</i>	
<i>service</i>	<i>float</i>	
<i>flag</i>	<i>float</i>	
<i>src_bytes</i>	<i>float</i>	
<i>dst_bytes</i>	<i>float</i>	
<i>hot</i>	<i>float</i>	
<i>logged_in</i>	<i>float</i>	
<i>num_compromised</i>	<i>float</i>	
<i>num_root</i>	<i>float</i>	
<i>is_guest_login</i>	<i>float</i>	
<i>Count</i>	<i>float</i>	
<i>srv_count</i>	<i>float</i>	
<i>error_rate</i>	<i>float</i>	
<i>srv_error_rate</i>	<i>float</i>	
<i>error_rate</i>	<i>float</i>	
<i>rv_error_rate</i>	<i>float</i>	
<i>same_srv_rate</i>	<i>float</i>	
<i>diff_srv_rate</i>	<i>float</i>	
<i>rv_diff_host_rate</i>	<i>float</i>	
<i>dst_host_srv_count</i>	<i>float</i>	
<i>dst_host_same_srv_rate</i>	<i>float</i>	
<i>dst_host_diff_srv_rate</i>	<i>float</i>	

[illegible]

- [illegible]

[illegible][illegible][illegible][illegible][illegible][illegible][illegible][illegible]

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3. Nama : Hasil
 Deskripsi : Berisi data hasil pengujian
 Primary Key : id

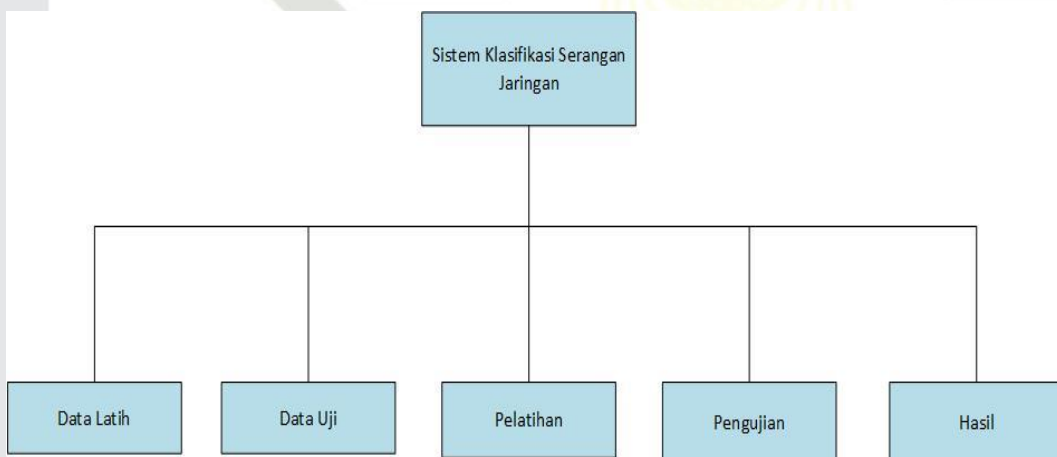
Berikut ialah detail tabel perancangan *database* data hasil pada Tabel 4.22 berikut:

Tabel 4.22 Tabel Data Hasil

Atribut	Type	Keterangan
id	Int(6)	Primary key
waktu	Text	
jumlah_data_latih	Int(11)	
jumlah_data_uji	Int(11)	
K	Int(6)	
confusion_matrix	Text	
accuracy_cm	double	
csv	Varchar(50)	

4.4.2 Perancangan Struktur Menu

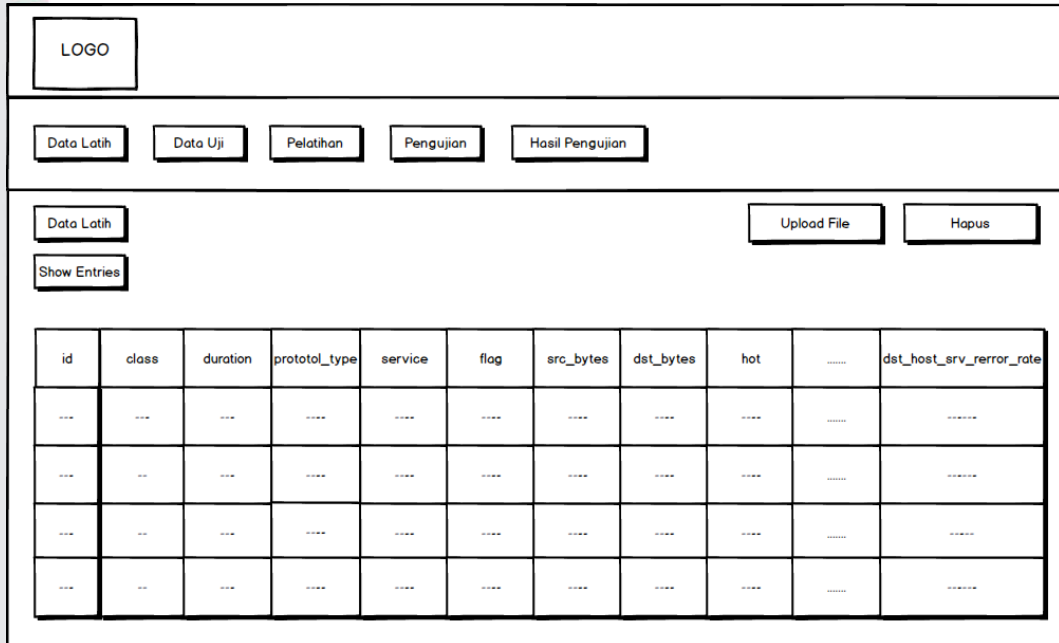
Perancangan struktur menu menggambarkan fitur-fitur atau menu yang terdapat pada sistem yang akan dibangun.. Struktur menu sistem yang akan dibangun dijabarkan pada Gambar 4.18 berikut:



Gambar 4.18 Struktur Menu Sistem

4.4.3 Perancangan Interface

Perancangan *interface* ini ialah perancangan tampilan antarmuka sistem yang akan dibangun berdasarkan menu yang telah dirancang sebelumnya. Hasil dari perancangan *interface* ini nantinya digunakan sebagai acuan ataupun dasar untuk



id	class	duration	prototol_type	service	flag	src_bytes	dst_bytes	hot	dst_host_srv_error_rate
---	---	---	----	----	----	----	----	----	-----	-----
---	---	---	----	----	----	----	----	----	-----	-----
---	---	---	----	----	----	----	----	----	-----	-----
---	---	---	----	----	----	----	----	----	-----	-----

Gambar 4.19 Interface Data Latih

2. Perancangan Menu Data Uji

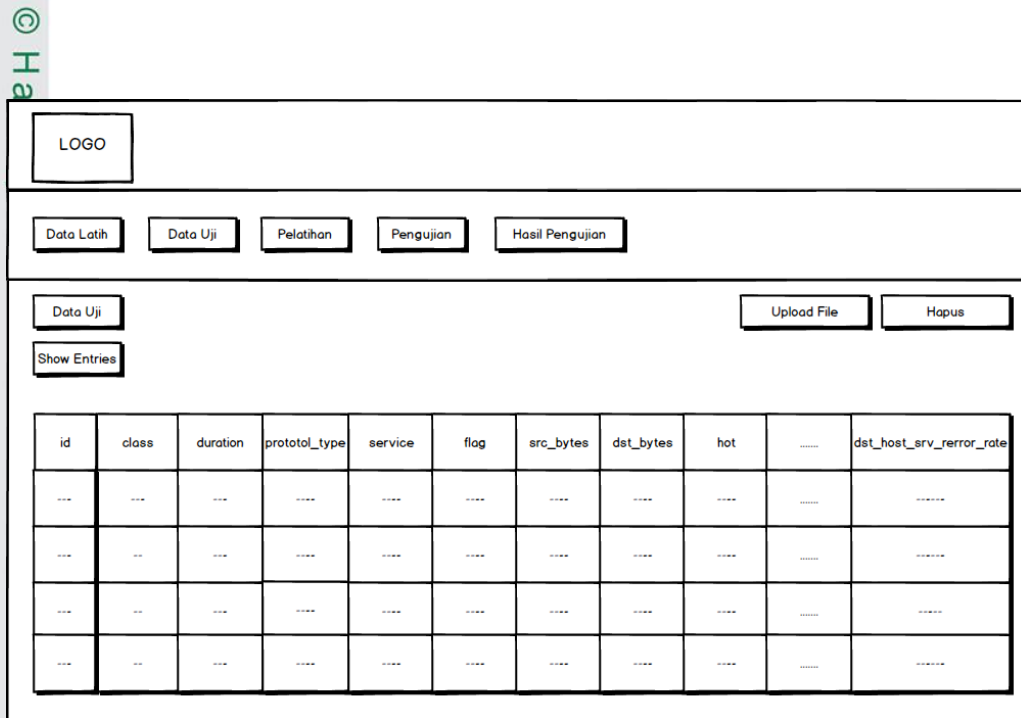
Berikut perancangan antarmuka menu data uji sistem pada Gambar 4.20:

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

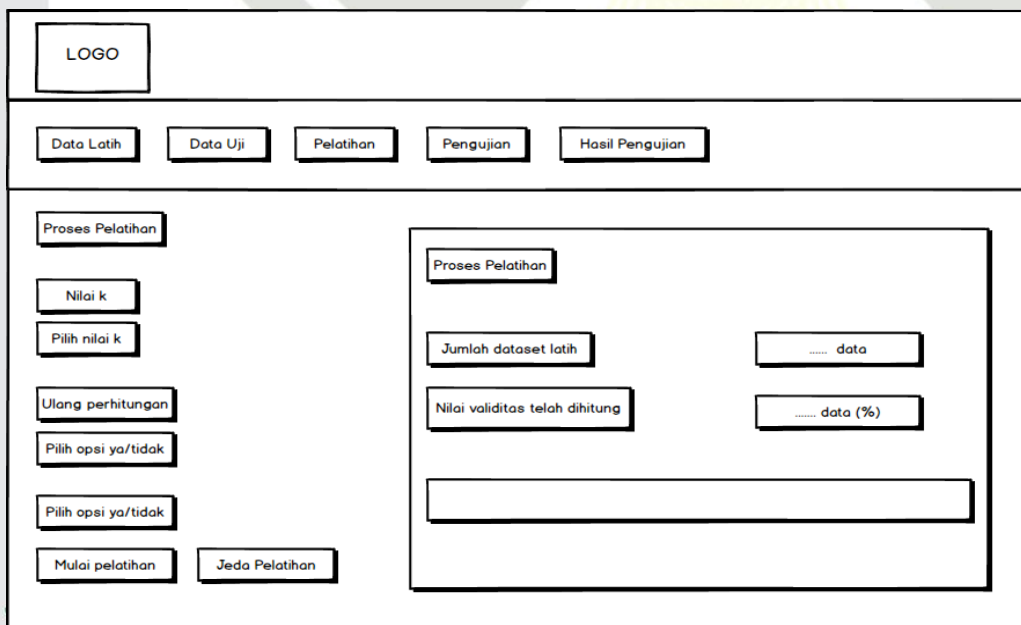


id	class	duration	prototol_type	service	flog	src_bytes	dst_bytes	hot	dst_host_srv_error_rate
...
...
...
...

Gambar 4.20 Interface Data Uji

3. Perancangan Menu Pelatihan

Berikut perancangan antarmuka menu pelatihan sistem pada Gambar 4.21:



Gambar 4.21 Interface Pelatihan

4. Perancangan Menu Pengujian

Berikut perancangan antarmuka menu pengujian sistem pada Gambar 4.22:

	dos	normal	probe	r2l	u2r
dos	0	0	0	0	0
normal	0	0	0	0	0
probe	0	0	0	0	0
r2l	0	0	0	0	0
u2r	0	0	0	0	0

Gambar 4.22 Interface Pengujian

5. Perancangan Menu Hasil

Berikut perancangan antarmuka menu data uji sistem pada Gambar 4.23:

No	Waktu Pengujian	Jumlah data latih	Jumlah data uji	Nilai k	Akurasi	Detail
--	----	---	---	---	---	-----

Gambar 4.23 Interface Hasil

- Hak Cipta Dilindungi Undang-Undang**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Kemudian gambar berikut ialah detail hasil pengujian dari masing-masing nilai K pada Gambar 4.24:

LOGO

Data Latih

Data Uji

Pelatihan

Pengujian

Hasil Pengujian

Detail Pengujian

Waktu Mulai Pengujian	Waktu Selesai Pengujian	Jumlah Data Latih	Jumlah Data Uji	K	Akurasi
----	---	---	---	---	---

Confusion Matrix

	dos	normal	probe	r2l	u2r
dos	0	0	0	0	0
normal	0	0	0	0	0
probe	0	0	0	0	0
r2l	0	0	0	0	0
u2r	0	0	0	0	0

Terklasifikasi Benar	-----
Terklasifikasi Salah	-----
CSV Hasil Pengujian	-----

Gambar 4.24 Interface Detail Hasil Pengujian



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB VI PENUTUP

1.1 Kesimpulan

Berdasarkan dari tahapan penelitian yang dilakukan mengenai klasifikasi serangan jaringan pada serangan jaringan dengan menggunakan metode MKNN diperoleh kesimpulan:

1. Metode MKNN dapat diterapkan untuk klasifikasi serangan jaringan pada data NSL-KDD. Klasifikasi menghasilkan *output* 5 kelas yakni Normal, Dos, U2R, R2L dan *Probe*.
2. Akurasi tertinggi yang diperoleh pada penelitian ini berdasarkan pengujian *confusion matrix* ialah 86,31% pada nilai $k=7$.

1.2 Saran

Berdasarkan tahapan serta hasil yang telah diperoleh pada penelitian ini, saran yang dapat diberikan guna untuk pengembangan penelitian selanjutnya :

1. Mengkombinasikan dengan menggunakan metode optimasi seperti algoritma genetika, tabu dan *particle swam optimization*.
2. Mengkombinasi parameter ataupun variabel pada data NSL-KDD. Tentukan parameter yang paling berpengaruh terhadap pengujian akurasi. Diharapkan dengan mengkombinasikan metode MKNN terhadap metode optimasi serta mengkombinasi parameter dengan menentukan parameter yang paling berpengaruh terhadap pengujian akurasi, dapat meningkatkan akurasi terhadap penelitian selanjutnya.

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR PUSTAKA

- Alder, R. (2004). *Snort 2.1 Intrusion Detection, Second Edition*. Rockland, MA 02370: Syngress Publishing, Inc.
- APJII. (2017). Penetrasi dan perilaku pengguna internet indonesia.
- Ariyus, D. (2006). *Internet Firewall*. Yogyakarta: Graha Ilmu.
- Dhanabal, L., dan Shantharajah, S. P. (2015). A Study on NSL-KDD Dataset for Intrusion Detection System Based on Classification Algorithms. *International Journal of Advanced Research in Computer and Communication Engineering*, 4(6), 446–452. <https://doi.org/10.17148/IJARCCCE.2015.4696>
- Fernanda, S. I., Ratnawati, D. E., dan Adikara, P. P. (2017). Identifikasi Penyakit Diabetes Mellitus Menggunakan Metode Modified KNearest Neighbor (MKNN).
- Ghozali, I. (2006). *Aplikasi Analisis Multivariate dengan Program SPSS*. Semarang: Universitas Diponegoro.
- Han, J., Kamber, M., dan Pei, J. (2012). *Data Mining Concepts and Techniques Third Edition*. Waltham: Morgan Kaufmann.
- Ibrahim, L. M., Basheer, D. T., dan Mahmod, M. S. (2013). A comparison study for intrusion database (KDD99 , NSL-KDD) based on self organization map (SOM) artificial neural network. *Journal of Engineering Science and Technology*, (January).
- Legre, B., dan Yadav, A. (2015). Performance analysis of NSL-KDD dataset using ANN. *International Conference on Signal Processing and Communication Engineering Systems - Proceedings of SPACES 2015, in Association with IEEE*, 92–96. <https://doi.org/10.1109/SPACES.2015.7058223>
- Larose, D. T. (2005). *Discovering Knowledge In Data An Introduction to Data Mining*. New Jersey: John Wiley dan Sons, Inc.
- Liao, H. J., Richard Lin, C. H., Lin, Y. C., dan Tung, K. Y. (2013). Intrusion detection system: A comprehensive review. *Journal of Network and Computer Applications*, 36(1), 16–24. <https://doi.org/10.1016/j.jnca.2012.09.004>
- Mas'ud Zubairi, A. (2014). *Analisis Klasifikasi Serangan Denial of Service (DOS)*

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Dengan Metode Decision Tree Menggunakan WEKA (Waikato Environment for Knowledge Analysis). Universitas Pembangunan Nasional.

Mongkareng, D., Setiawan, N. A., dan Permanasari, A. E. (2017). Implementasi Data Mining dengan Seleksi Fitur untuk Klasifikasi Serangan pada Intrusion Detection System (IDS). *Citee*, 314–321.

Mutrofin, S., Kurniawardhani, A., Izzah, A., dan Masrur, M. (2014). Optimasi teknik klasifikasi modified k nearest neighbor menggunakan algoritma genetika. *Jurnal Gamma*, (September), 130–134.

Mydza, D. M. (2011). *Analisa Dan Konfigurasi Network Intrusion Prevention System (Nips) Pada Linux Ubuntu 10 . 04 Lts Analisa Dan Konfigurasi Network Intrusion Prevention System (Nips)*. State Islamic University of Sultan Syarif Kasim Riau.

Okfalisa, Mustakim, Gazalba, I., dan Reza, N. G. I. (2017). Comparative Analysis of K-Nearest Neighbor and Modified K-Nearest Neighbor Algorithm for Data Classification. *IEEE*, 294–298.

Parvin, H., Alizadeh, H., dan Minaei-bidgoli, B. (2008). MKNN : Modified K-Nearest Neighbor. *Proceedings of the World Congress on Engineering and Computer Science, WCECS*, 22–25. <https://doi.org/10.1.1.149.545>

Parvin, H., Alizadeh, H., dan Minati, B. (2010). A Modification on K-Nearest Neighbor Classifier. *Global Journal of Computer Science and Technology*, 10(14), 37–41.

Rasetyo, E. (2012). Fuzzy K-Nearest Neighbor In Every Class Untuk Klasifikasi Data. *Seminar Nasional Teknik Informatika*, (Santika), 57–60.

Simanjutak, T. H., Mahmudy, W. F., dan Sutrisno. (2017). Implementasi Modified K-Nearest Neighbor Dengan Otomatisasi Nilai K Pada Pengklasifikasian Penyakit Tanaman Kedelai.

Soewarno. (1995). *Hidrologi (Aplikasi Metode Statistik untuk Analisa Data) Jilid 1*. Bandung: Nova.

Tavallae, M., Bagheri, E., Lu, W., dan Ghorbani, A. A. (2009). A detailed analysis of the KDD CUP 99 data set. *IEEE Symposium on Computational Intelligence for Security and Defense Applications, CISDA 2009*, (Cisda), 1–6.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

<https://doi.org/10.1109/CISDA.2009.5356528>

Wafiyah, F., Hidayat, N., dan Perdana, R. S. (2017). Implementasi Algoritma Modified K-Nearest Neighbor (MKNN) untuk Klasifikasi Penyakit Demam.

Wilson, D. R., dan Martinez, T. R. (1997). Improved heterogeneous distance functions. *Journal of Artificial Intelligence Research*, 6, 1–34. <https://doi.org/10.1613/jair.346>

Wirawan, I. N. T., dan Eksistyanto, I. (2015). Penerapan Naive Bayes Pada Intrusion Detection System Dengan Diskritisasi Variabel. *Jurnal Ilmiah Teknologi Informasi*, 13, 182–189.

Yang, Q., dan Wu, X. (2006). 10 Challenging Problems in Data Mining Research. *International Journal of Information Technology dan Decision Making*, 05(04), 597–604. <https://doi.org/10.1142/S0219622006002258>

UIN SUSKA RIAU

DAFTAR RIWAYAT HIDUP

Informasi Personal



Nama	: FAKHRIAL IRSYADI
Tempat, Tanggal Lahir	: Pekanbaru, 05 Juni 1995
Status Pernikahan	: Belum Menikah
Agama	: Islam
Anak Ke-	: 1
Jumlah Sdr.	: 4
Tinggi Badan	: 165 cm
Jenis Kelamin	: Laki - laki

Informasi Domisili

Kebangsaan	: Indonesia
Alamat Sekarang	: Jl. Bayur Raya C 21 No 1 Perumahan Pandau Permai
Nomor HP	: 082268375513
Email	: fakhrial.irsyadi@students.uin-suska.ac.id

Informasi Pendidikan

1. Tahun 2000-2001	: TK YKWI Riau
2. Tahun 2001-2003	: SD Negeri 025 Siak Hulu, Riau
3. Tahun 2007-2008	: SMP Negeri 25 Pekanbaru, Riau
4. Tahun 2010-2013	: SMK Negeri 2 Pekanbaru, Riau
5. Tahun 2013-2019	: Tekni Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.